(E80-10324) IRRIGATED LANDS ASSESSMENT FOR WATER MANAGEMENT APPLICATIONS PILOT TEST (APT) Final Report (California Univ.) 156 p HC A08/MF A01 CSCL 08H N80-32815
Unclas
G3/43 00324

SPACE
SCIENCES
LABORATORY

IRRIGATED LANDS ASSESSMENT FOR WATER MANAGEMENT APPLICATIONS PILOT TEST (APT)

80-10324
CR-163404

Principal Investigator: Robert N. Colwell

Remote Sensing Research Program
Berkeley Campus
Co-Experimenter,
Project Manager: Sharon L. Wall
Contributors: Catherine E. Brown
Kevin J. Dummer
Thomas W. Gossard
Randall W. Thomas

Geography Remote Sensing Unit
Santa Barbara Campus
Co-Experimenters: John E. Estes
Larry Tinney
Project Manager: Jay Baggett

Prepared for National Aeronautics and Space Administration
Ames Research Center, Moffett Field, California under
NASA Grant NSG 2207

Final Report
1 January 1977 - 31 January 1980
Space Sciences Laboratory
Series 21, Issue 5

UNIVERSITY OF CALIFORNIA, BERKELEY

IRRIGATED LANDS ASSESSMENT FOR WATER MANAGEMENT
APPLICATIONS PILOT TEST (APT)

Principal Investigator: Robert N. Colwell

Remote Sensing Research Program
260 Space Sciences Laboratory
University of California/Berkeley, CA 94720

Co-Experimenter and Project Manager: Sharon L. Wall
Contributors: Catherine E. Brown
Kevin J. Dummer
Thomas W. Gossard
Randall W. Thomas

Geography Remote Sensing Unit
University of California/Santa Barbara, CA 93106

Co-Experimenters: John E. Estes
Larry R. Tinney
Project Manager: Jay Baggett

Prepared for National Aeronautics and Space Administration
Ames Research Center, Moffett Field, California
NASA Grant NSG 2207

Final Report
1 January 1977 - 31 January 1980
Space Sciences Laboratory
Series 21, Issue 5
University of California, Berkeley

NASA TECHNICAL OFFICER AND RESEARCH SUPPORT

Ethel H. Bauer
Technology Applications Branch
M/S 242-4
NASA/Ames Research Center
Moffett Field, CA 94035

The research reported on under this grant (NSG 2207) was conducted previously under NASA Contract NAS 5-20969 (Goddard Space Flight Center, Greenbelt, MD). Research is continuing under NASA Cooperative Agreement NCC 2-54 (Ames Research Center, Moffett Field, CA).

ACKNOWLEDGMENTS

Much of the success of the project is directly attributable to continuing and increasing interest and support from the California Department of Water Resources. We would especially like to acknowledge Mr. Glenn Sawyer and Mr. James Wardlow (State Offices, Sacramento); Mr. J. Kono, Mr. F. Littel and Mr. G. Sato (Central District, Sacramento); Mr. C. Ferchaud, Mr. R. McGill and Mr. G. Pixley (Northern District, Red Bluff); Mr. J. Berthelot, Mr. F. Stumpf and Mr. J. Williams (San Joaquin District, Fresno); and Mr. R. Smith and Mr. R. Yoha (Southern District, Los Angeles), who have been particularly active participants on the project.

In addition, we would like to express our appreciation of our colleagues within the University of California: from the Remote Sensing Research Program (Berkeley Campus), the Director of the Program, Mr. Stephen DeGloria; engineer, Mr. Anthony Travlos; photographers, Mr. Stephen Harui and Ms. Dorothy Robinson; programmers, Ms. Evelyn Gray, Mr. Agnis Kaugars, Mr. Paul Ritter, Mr. Derek Taylor and Mr. George Vasick; and staff research personnel, Dr. Steven Daus, Mr. Charles Henderson and Mr. Vincent Vesterby; from the Geography Remote Sensing Unit (Santa Barbara campus), photographer, Mr. Joseph Scepan; programmer, Mr. Wayne Hallada; and staff research personnel, Mr. David Ferguson, Ms. Anne Meshkov and Ms. Tara Torburn.

In the course of our work we have also received valuable assistance from Mr. Florin Core and Mr. Mike Rector of the Kern County Water Agency; and from Mr. Bryce Jope and Mr. George Katibah of CALTRANS (Sacramento).

TABLE OF CONTENTS

Appendicies

1.00 INTRODUCTION*

With the cultivation of some 200 commercial crops, California's agriculture is one of the most diversified in the world. The State leads the nation in the production of 47 commercial crop and livestock commodities and is one of the top five producers of an additional 19. Gross cash receipts from farm marketings in 1978 totaled $10.4 billion. With this income, California continues as the leading farm state with nearly 10 percent of the nation's cash receipts.

This abundance of agricultural output results from the cultivation of approximately 13.3 million hectares (32.8 million A, 1978). The combined acreage of principal crops in 1978 totaled 3.8 million hectares (9.4 million A), a 5 percent increase in harvested area since 1977. Field crops (2.7 million hectares, 6.7 million A), fruit and nut crops (.7 million hectares, 1.7 million A) and vegetable and melon crops (.4 million hectares, .9 million A) yielded 43.1 million tons of harvested farm products.[1] Much of the success of this agricultural production is founded on the availability of water for irrigation. The California Department of Water Resources (DWR) estimates that approximately 3.8 million hectares (9.5 million A) are irrigated at least once during the growing season. This water is derived from surface sources, groundwater extraction and the construction of large-scale water transport projects. Agriculture is the prime recipient of the available water, utilizing about 85% of the supply.

In 1957, California Water Code Section 10005 established the California Water Plan. It is a "comprehensive master plan to guide and coordinate the planning and construction of works required for the control, protection, conservation and distribution of the water of California to meet present and future needs for all beneficial uses and purposes in all areas of the State".[2] The responsibility for updating and supplementing the Plan was assigned to the Department of Water Resources.

"The Department carries out this responsibility through a statewide planning program, which guides the selection of the most favorable pattern for the use of the State's water resources, considering all reasonable alternative courses of action. Such alternatives are evaluated on the

---

* All principal measurements and calculations were performed using customary units.

1 Department of Food and Agriculture, State of California, "California Principal Crop and Livestock Commodities - 1978"

2 Department of Water Resources, State of California, "The California Water Plan Outlook in 1974, "Bulletin No. 160-74, November, 1974

basis of technical feasibility and economic, social, and institutional factors. The program comprises:

- Periodic reassessment of existing and future demands for water for all uses in the hydrologic study areas of California.

- Periodic reassessment of local water resources, water uses, and the magnitude and timing of the need for additional water supplies that cannot be supplied locally.

- Appraisal of various alternative sources of ground water, surface water, reclaimed waste water, desalting, geothermal resources, etc. - to meet future demands in the areas of water deficiency.

- Determination of the need for protection and preservation of water in keeping with protection and enhancement of the environment.

- Evaluation of water development plans.[3]

A summary status of conditions and expectations is published every four years in the form of a comprehensive bulletin (Bulletin 160) that is used to provide information to aid in guiding and coordinating the use of California's water resources.

To meet these responsibilities, DWR has long recognized the need for specific land use data as an input to state water planning. Since the late 1940's the Department has been performing a continuing survey to monitor land use changes over the state. Because of manpower and budgetary constraints, only a portion of the state (approximately one-seventh) is surveyed during any given year. In DWR's surveys, two types of output are produced, (1) land use surveys which record the nature and extent of present water-related land development, and (2) land classification surveys designed to determine the location and extent of lands with physical characteristics suited to specific kinds of development. The more pertinent of these surveys to the projects discussed in this report, is the land use survey. It is compiled through the interpretation of current 35 mm aerial photography supplemented with field inspections. Tabulations of the acreage of each specific land use class are then summarized by 7-1/2 minute quad sheet, county and other area subdivisions such as water agency or hydrographic area. Figures 1-1 and 1-2 show the land use legend and a completed land use map prepared by DWR.

As seen in Figures 1-1 and 1-2, each parcel of agricultural land has been designated as either irrigated, the prefix "i," or non-irrigated, "n". This condition is determined by the interpretation of aerial photography and the gathering of supplementary field data as mentioned above. From the

---

[3] Ibid

data collected, DWR is able to generate maps showing the land use classification to cover type, including crop identification, and the acreage of irrigated lands. Since each land use is associated with a specific water demand, total water consumption forecasts can then be made. Due to the limitations of the one date survey, however, the DWR survey is not considered accurate as to the proportion of acreage devoted to small grains or multiple cropping.

California receives an annual average of 200 million acre-feet of precipitation. Most of the runoff, approximately 70,000,000 acre-feet, occurs in areas with the lowest population densities. As a result large-scale water systems, both state and federal, have been constructed to store and transport water from the areas of accumulation to the areas of demand. In recent years California like much of the West had been experiencing a major drought. In normal years approximately 85% of the total water used would be consumed by agriculture. In 1977, state officials were expecting a 10 million acre-foot deficit. Because of this, state and federal water managers initiated stringent reductions in water deliveries. The Bureau of Reclamation reduced its deliveries to less than half of their usual Central Valley Project (Federal jurisdiction) allotments. Likewise, the State Water Project (State jurisdiction) was forced to curtail water deliveries to its 24 contracting districts to 1.8 million acre-feet, down 1.6 million acre-feet from projected demand. Since California alone supplies approximately 40% of the nation's summer fruit and vegetable crops, the loss in production caused by the drought impacts agribusiness and price structures nationwide.

The drought dramatically emphasized the need for accurate and timely information on the extent of irrigation and the nature of agriculture as input for water management decisions. In addition to their normal survey techniques, DWR has been actively participating since 1975 with NASA and the University of California on several projects designed to investigate the feasibility of estimating irrigated acreage and determining cropping practices within the state utilizing a Landsat-based remote sensing system. Based on the results of these studies, information acquired from the analysis of satellite imagery may become a valuable supplement to the land use information presently collected by DWR. The use of the satellite system allows DWR the opportunity to analyze data from several dates during the growing season and the ability to collect data over the entire state in one year.

AGRICULTURE

Each parcel of agricultural land use is labeled with a notation consisting basically of three symbols. The first of these is a lower case "i" or "n" indicating whether the parcel is irrigated or nonirrigated. This is followed by a capital letter and number which denote the use group and specific use as shown below.

C SUBTROPICAL FRUITS
1 Grapefruit
2 Lemons
3 Oranges
4 Dates
5 Avocados
6 Olives
7 Miscellaneous subtropical fruits

D DECIDUOUS FRUITS AND NUTS
1 Apples
2 Apricots
3 Cherries
5 Peaches and Nectarines
6 Pears
7 Plums
8 Prunes
9 Figs
10 Miscellaneous or mixed deciduous
12 Almonds
13 Walnuts

G GRAIN AND HAY CROPS
1 Barley
2 Wheat
3 Oats
6 Miscellaneous and mixed hay and grain

F FIELD CROPS
1 Cotton
2 Safflower
3 Flax
4 Hops
5 Sugar beets
6 Corn (field or sweet)
7 Grain sorghums
8 Sudan
9 Castor beans
10 Beans (dry)
11 Miscellaneous field

T TRUCK AND BERRY CROPS
1 Artichokes
2 Asparagus
3 Beans (green)
4 Cole crops
6 Carrots
7 Celery
8 Lettuce (all types)
9 Melons, squash, and cucumbers (all kinds)
10 Onions and garlic
11 Peas
12 Potatoes
13 Sweet potatoes
14 Spinach
15 Tomatoes
16 Flowers and nursery
18 Miscellaneous truck
19 Bushberries
20 Strawberries
21 Peppers (all types)

P PASTURE
1 Alfalfa and alfalfa mixtures
2 Clover
3 Mixed pasture
4 Native pasture

V VINEYARDS

R RICE

I IDLE
1 Land cropped within the past three years but not tilled at time of survey
2 New lands being prepared for crop production

S SEMIAGRICULTURAL AND INCIDENTAL TO AGRICULTURE
1 Farmsteads
2 Feed lots (livestock and poultry)
3 Dairies
4 Lawn areas

Special conditions are indicated by the following additional symbols and combinations of symbols.

A ABANDONED ORCHARDS AND VINEYARDS
F FALLOW (tilled but not cropped at time of survey)
S SEED CROPS
Y YOUNG ORCHARDS AND VINEYARDS
X PARTIALLY IRRIGATED CROPS

INTERCROPPING (or interplanting) is indicated as follows: ; $\frac{D13\text{-}y}{T9}$ = a melon crop planted between rows of young walnut trees

URBAN

UC - URBAN COMMERCIAL
UC 1 Miscellaneous establishments (offices and retailers)
UC 2 Hotels
UC 3 Motels
UC 4 Apartments, barracks (three family units and larger)
UC 5 Institutions (hospitals, prisons, reformatories, asylums, etc., having a reasonably/stable 24-resident population)
UC 6 Schools (yards mapped separately if large enough)
UC 7 Municipal auditoriums, theaters, churches, buildings, and stands associated with race tracks, football stadiums, baseball parks, rodeo arenas, etc.
UC 8 Miscellaneous high water use (indicates a high water use not covered above)

UI - URBAN INDUSTRIAL
UI 1 Manufacturing, assembling, and general processing
UI 2 Extractive industries (oil fields, rock quarries, gravel pits, public dumps, rock and gravel processing plants, etc.)
UI 3 Storage and distribution (warehouses, substations, railroad marshalling yards, tank farms, etc.)
UI 6 Saw mills
UI 7 Oil refineries
UI 8 Paper mills
UI 9 Meat packing plants
UI 10 Steel and aluminum mills
UI 11 Fruit and vegetable canneries and general food processing
UI 12 Miscellaneous high water use (indicates a high water use not covered above)

UV - URBAN VACANT
UV 1 Miscellaneous unpaved areas
UV 4 Miscellaneous paved areas

UR - URBAN RESIDENTIAL
One and two family units, including trailer courts

RECREATION

RR RESIDENTIAL
Permanent and summer home tracts within a primarily recreational area. (The estimated number of houses per acre is indicated by a number in the symbol.)

RC COMMERCIAL
Commercial areas within a primarily recreational area (includes motels, resorts, hotels, stores, etc.)

RT CAMP AND TRAILER SITES
Camp and trailer sites in a primarily recreational area

P PARKS

NATIVE

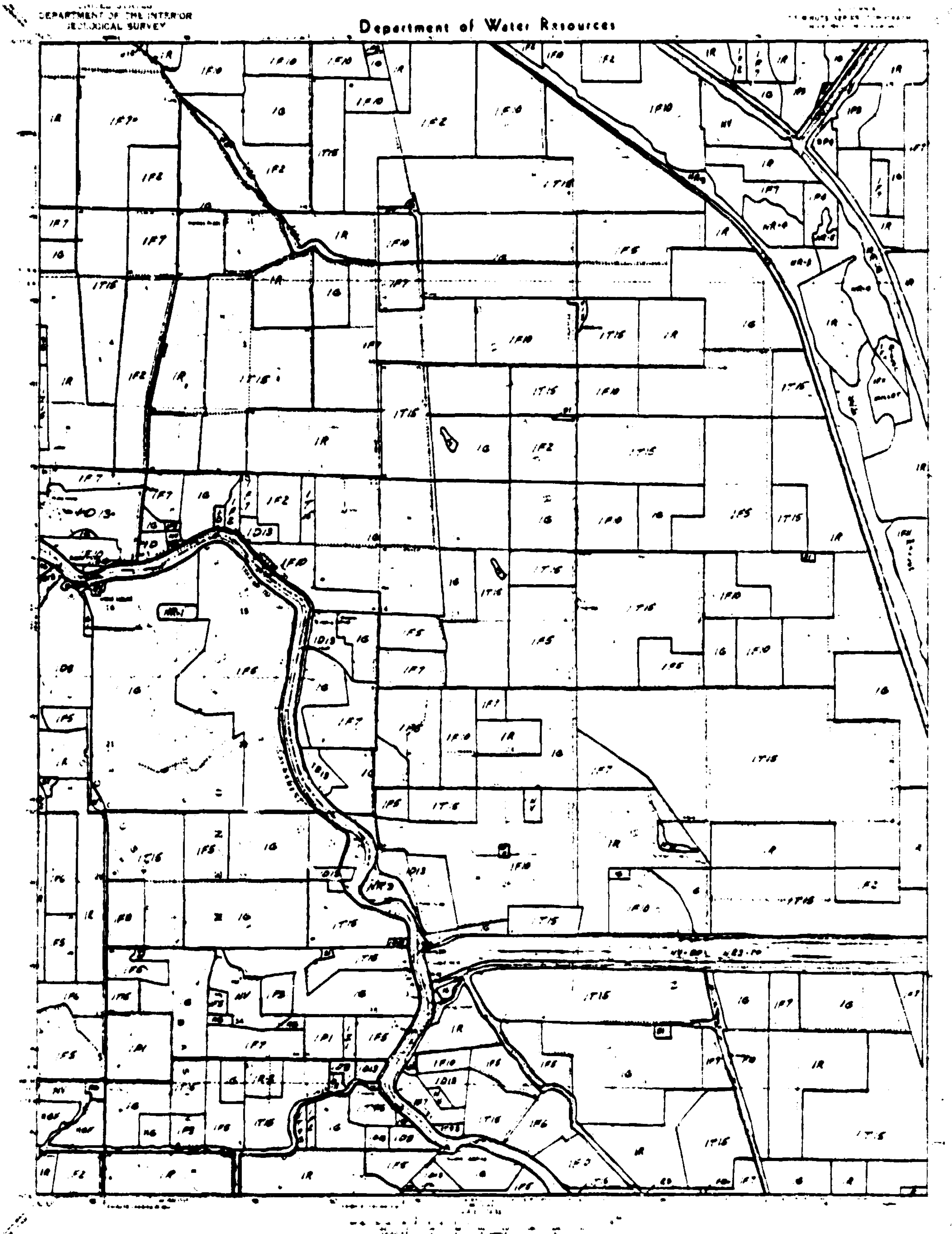
NV NATIVE VEGETATION

NR RIPARIAN VEGETATION
NR 1 Swamps and marshes
NR 2 Meadowland

NW WATER SURFACE

NC NATIVE CLASSES UNSEGREGATED

Figure 1-1. Legend developed by the California Department of Water Resources and used in their land use surveys.



1976
LAND USE

TISDALE WEIR CALIF

Figure 1-2. An example of a completed 7-1/2 minute quadrangle land use survey map prepared by DWR.

2.0 SUMMARY OF RESULTS: 15 APRIL 1975 - 31 DECEMBER 1978

Since 1975, the California Department of Water Resources has been cooperating with NASA and the University of California on a series of projects designed to address the applicability of satellite data as input to water management decisions. Research and demonstration have been supported by: 1) NASA Contract NAS 5-20969 (Goddard Space Flight Center) and 2) NASA Grant NSG 2207 (Ames Research Center). Although the majority of the described work in this final report was supported by Ames Research Center, the results are based on preliminary work developed under contract with Goddard Space Flight Center. Because each succeeding year's work has been based on accomplishments of the previous years, a description of the methodologies, rationales and results of the initial contract work are included. Following that, summaries of the first two years of this grant are given. The remainder of the final report details the results of this year's (1979) effort. Further work on the project is continuing under NASA Cooperative Agreement NCC 2-54 (Ames Research Center). Figure 2-1 diagrams the stepwise support of the project.

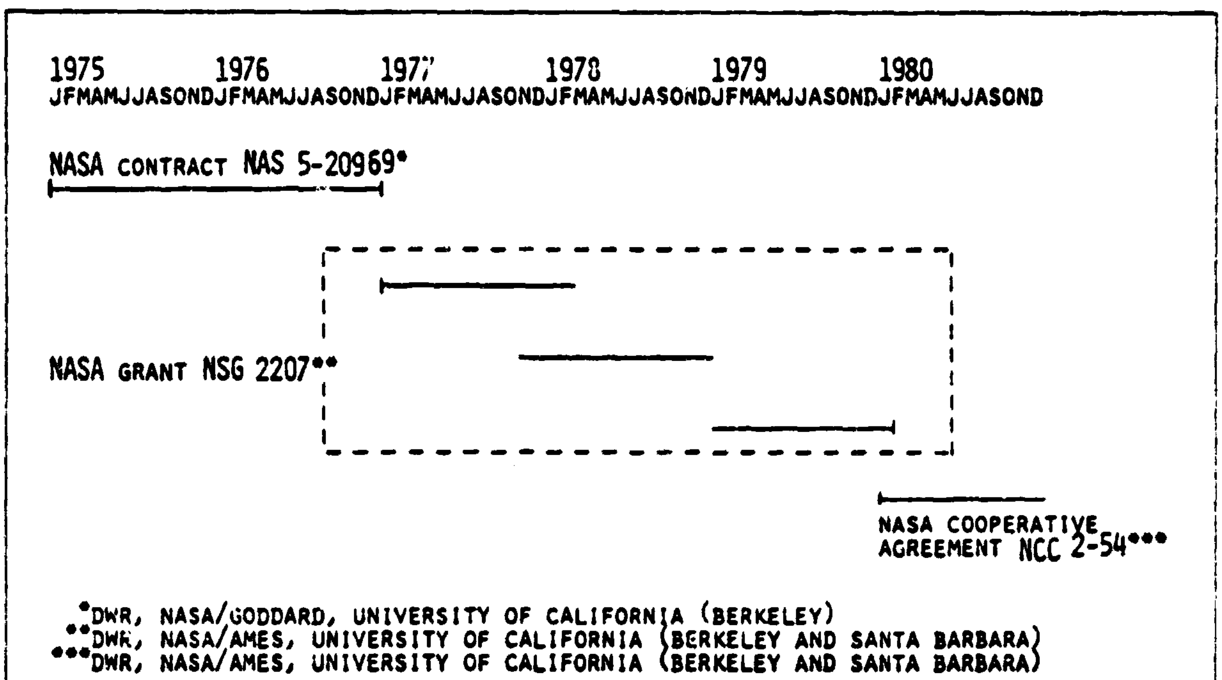


Figure 2-1. Work on the estimation of irrigated land in California has been supported by NASA since 1975. This figure diagrams the timeframes of the three funding vehicles. NASA Grant 2207 provided the support for the work covered by this final report.

2.1 AN INVENTORY OF IRRIGATED LANDS FOR SELECTED COUNTIES WITHIN THE STATE OF CALIFORNIA BASED ON LANDSAT AND SUPPORTING AIRCRAFT DATA (15 April 1975 - 15 January 1977)

The first cooperative effort between NASA/GSFC, DWR and the Remote Sensing Research Program of U.C. Berkeley began on the 15th of April 1975 and continued until January 15, 1977. The three main objectives of the study were: (1) to develop a process for providing irrigated acreage on a regional basis using Landsat; (2) to develop a technique that would provide this estimate in one year, and (3) to achieve a level of precision for the State to within $\pm 3\%$ at the 99% level of confidence.

Selected in conjunction with DWR, ten counties representing much of the agricultural diversity found in California were defined as the study area. Seven of the ten counties were located in the Central Valley; others were located in coastal and mountain areas. After exclusion areas had been removed, the total population subject to sampling and interpretation was approximately 1,500,000 hectares (3,707,000 A). Exclusion areas were defined as areas not subject to irrigation (urban, wildland, wildlife refuges) and areas where information on irrigation was so good as to make sampling unnecessary (established orchards).

A three phase sample design based on a sampling frame of area units with stratification by county was used. The three phase design was selected to maximize the advantages of spectral reflectance and field pattern (auxiliary variable data) available on Landsat and aerial photography as they relate to irrigated acreage. Multiple dates of Landsat were used as Phase I to provide relatively inexpensive, county-wide estimates of irrigated proportion. Multitemporal vertical color aerial photography, used as Phase II, provided a cost-effective means to correct the Landsat estimates for bias. Finally measurements made on a small sample of Phase III ground units were used, in turn, to calibrate the aerial photography estimates and provide the most accurate information on crop type and irrigation.

A rectangular sample grid of 1.6 by 8.0 kilometer (1 x 5 miles) sample units was defined to cover each county. Since no prior irrigated acreage variance versus sample unit dimension data was available, sample unit size and shape were chosen based on practical considerations. These considerations dealt with ease of data acquisition and measurement at each sample stage.

In order to determine Phase I, II and III sample sizes by county (stratum) that would be expected to support the statewide $\pm$ 3%, 99% level of confidence irrigated acreage precision goal, a preliminary population model was constructed. Sample size (number of sample units) allocations were based on previously published estimates of proportion of area irrigated by county, approximate between phase cost ratios and a non-linear programming algorithm which minimizes cost, subject to constraints on variance. Samples were allocated with equal probability at each sample

phase within each county. Sample units eligible for selection were confined to those selected for measurement at the previous phase. For the entire ten county test site, 1292 Phase I, 90 Phase II and 18 Phase III units were selected.

A three phase regression estimation system was chosen to provide irrigated acreage proportion estimates. This model was thought to represent the between phase proportion relationships most accurately. The mean and variance estimators followed the treatment given by Tikkiwal (1955 and 1967). Basically, the estimators were iterative such that the Phase III (ground) estimator used the Phase II (aerial photo) estimator which in turn used the Phase I (Landsat) estimator.

The multitemporal capabilities available with Landsat offer obvious benefits for monitoring an agricultural growing season. Three time periods were selected for analysis: (1) June - to monitor small grains and establish a base for multiple cropping, (2) August - to provide data on maximum canopy coverage expected for many irrigated crops, and (3) September - to continue multiple cropping observations. Interpretation for Phase I was done on Multidate Landsat mosaics of each country that had been enlarged to 1:154,000. The August imagery acted as the base date for two major reasons. First, it is the height of the growing season when maximum vegetation cover is present. And second, in nearly all of the agricultural areas of California if a crop is growing in August it can be safely assumed to be irrigated. The May and October date imagery is used to control early harvested crops and multiple cropped areas.

The multitemporal large scale color aerial photography used as Phase II was procured using a Twin Commanche aircraft, equipped with a vertical closed circuit TV system for location and a Nikon 35mm camera for photography. After enlargement to the standard 3R size (scale approximately 1:21,000) the photography was mosaicked into strips that covered each sample unit. Each sample unit was then interpreted to obtain an estimate of the irrigated area within it. Multitemporal ground data (Phase III) was also collected for a sub-set of the sample units flown with aerial photography.

The results of the interpretation and ground data collection were tabulated and input to a Fortran program, MPHASE, written at Berkeley and designed to calculate the multiphase estimate, variance, standard error, relative standard error and sample correlation coefficients for each county. Results were a regional estimate, summarized by county which calculated 80.17% (1,202,401 ha; 2,971,827 A) of the population estimated to be irrigated. The confidence interval of the estimate is shown below. Since the population sampled in this study represented less than half the agricultural land in California, a sample of the larger area would be expected to produce precision performance approaching ± 3% at the 99% level requested by DWR for statewide reporting.

Table 2-1. Confidence interval of the estimate of irrigated land in ten counties in California.

Confidence interval of the estimate (half width expressed as percent)		
$1 - \alpha = .68$	$1 - \alpha = .95$	$1 - \alpha = .99$
$t = 1.00$	$t = 1.98$	$t = 2.358$
$\pm$ 2.73	$\pm$ 5.41	$\pm$ 6.44

Evaluation by a University of California resource economist found that the costs of the inventory compared favorably with a hypothetical DWR-style survey of irrigated acreage only (approximately 3¢ hectare/1.2¢ acre). He further found that the results approximated comparable estimates produced by the Ag Census and county agricultural commissioners. In addition it was feasible to complete the project and generate the statistics within the DWR time requests.

With encouraging results from this first effort, a second project was undertaken. This cooperative study by NASA/ARC, DWR and the University of California (Berkeley and Santa Barbara campuses) was to continue the development of techniques to optimize the estimation of irrigated acreage and test on a large yardstick region. Additionally, work was initiated on the use of computer assisted analysis techniques for estimating irrigated acreage. Work on manual and computer assisted analysis techniques for determining specific crop types was also begun.

2.2 DETERMINING THE USEFULNESS OF REMOTE SENSING FOR ESTIMATING WATER DEMAND IN CALIFORNIA (1 January 1977 - 28 February 1978)

Two main test sites were selected for study: (1) the Sacramento Valley (1,977,000 hectares, 4,885,000 A) in northern California, and Kern County (404,700 hectares, 1,000,000 A) in the southern San Joaquin Valley. On both sites the DWR had performed 100% land use surveys.

In the Sacramento Valley Test site the proportion irrigated was estimated for the entire fourteen county region using manual analysis techniques developed in the previous project. To this end, a stratification recommended at the end of the Irrigated Lands Project was developed and produced for the region.

The main purposes of the stratification were to more optimally allocate sample units and control measurement error. The stratification developed was based on agricultural practices, environmental conditions and field size, the major factors affecting manual analysis performance. Six agricultural practice strata were identified. These strata were composed of areas that, on Landsat 1:1,000,000 color composite transparencies, appear to be:

Stratum Number	Stratum Description
1	Generally dry farmed
2	Field Crops - fields generally less than 16 hectares (40 A)
3	Field Crops - fields generally from 16 - 32 hectares (40 - 79A)
4	Field Crops - fields generally 33 hectares (80A) or greater
5	Orchards & vinyards - fields generally less than 16 hectares (40 A)
6	Orchards & vinyards - fields generally 16 hectares (40 A) or greater

The multiphase sampling design was maintained although some refinement was necessary; sample allocation was controlled by the strata and a two-phase rather than a three-phase sample design was used. Since the project was begun after the 1976 growing season, multitemporal aerial photography and ground data were not available. The photo and ground measurements were considered as a single phase (Phase II). The $\pm$ 3% at the 99% level of confidence for statewide estimation was continued. Based on this 1830 Phase I and 141 Phase II units were allocated.

Multitemporal Landsat from 30 May, 28 August and 3 October was used with interpretation and tabulation as before. The results of this study showed 54.24% (1,072,277 ha, 2,649,561 A) of the population estimated to be irrigated. A decrease in the relative error from ± 2.73% to ± 1.52% was achieved. The resulting confidence intervals are shown below:

Table 2-2. Confidence interval of the estimate of irrigated land in the fourteen-county Sacramento Valley Test Site.

Confidence interval of the estimate (half width expressed as percent)		
$1 - \alpha = .68$	$1 - \alpha = .95$	$1 - \alpha = .99$
$t = 1.00$	$t = 1.992$	$t = 2.631$
± 1.52	± 3.03	± 3.99

In addition a single county comparison was completed in which 21 individual 7-1/2 minute quadrangles were compared. Based on this comparison the Landsat measurement came within 4,047 hectares (10,000 A) of the 117,289 hectares (289,816 A) tabulated by DWR. Interpretation performed, on the same area by DWR personnel familiar with the county, came within 2,428 hectares (6,000 A) of the tabulated total.

In selected areas within the Sacramento Valley Site, computer assisted analysis techniques were tested for the estimation of irrigated land. Using multitemporal digital tapes, unsupervised and maximum likelihood classification techniques were used to estimate proportion irrigated on a single 7-1/2 minute quadrangle. Based on this classification, 67.52% of the area was estimated to be irrigated compared to 64.54% irrigated tabulated from the DWR land use survey.

In Kern County, a study to test the ability to <u>map</u> irrigated land using manual analysis techniques and multitemporal Landsat was also done. Work based on earlier studies with the Kern County Water Agency indicated that 95% of the area was correctly mapped as irrigated.

The second phase of this project was devoted to specific crop type estimation and mapping. Certain basic data was used in the work done on the various test sites which included: ancillary data such as historical crop acreages, trends, crop calendars; 100% aerial photography and ground data from DWR; regional multidate crop keys (Landsat); regional crop determination matrices; and multitemporal Landsat data in a variety of forms. A major task within this phase was the estimation of small grains within the entire Sacramento Valley Test Site and selected sites within Kern County. Using manual analysis techniques, multiphase sampling, stratification and multitemporal Landsat (19 March, 30 May, 26 June) techniques were developed to estimate the proportion of small grains within the Sacramento Site. The regional estimate was again summarized by county with 13.35% (264,005 ha, 652,348 A) estimated to be small grains. The relative error of this estimate was 6.26% or + 8.09% at the 90% level of confidence. Within Kern County, a mapping task to detect and map small grains was done on the Wheeler Ridge-Maricopa Test Site. Results of this task indicated that 85% of the small grains fields were correctly mapped.

Additional crop specific estimation and mapping was done for a variety of crops including alfalfa, sugar beets, cotton, tomatoes, safflower, melons, lettuce and fallow. Using multidate Landsat, per class accuracies averaged 71% in Kern County; when grouped in water consumptive use classes, accuracies increased to 84%. Safflower mapping in the Sacramento Valley Site averaged 92% correct; acreages, however, were very low in 1976.

A final task of the Agricultural Water Demand Project was initializing the definition of parameters for regionalizing the state by defining the varied agricultural regimes of California and by determining the typical signature of these varied areas based on photomorphic as well as cultural and physical factors. The final definition and production of the regionalization were completed during the first year of the Applications Pilot Test (APT) which began in November 1977 and ended December 1979.

2.3 IRRIGATED LANDS ASSESSMENT FOR WATER MANAGEMENT - APPLICATIONS PILOT TEST (APT) (1 November 1977 - 31 December 1978)

The first year of the APT (November 1977 - December 1978) focused on constructing a framework for a large scale demonstration and technology transfer. The specific objectives of the first year were to: (1) produce a general regionalization of the state (briefly described above), (2) extend the stratification, described above, to the San Joaquin Valley, (3) continue development and demonstration of techniques for DWR-defined interests in several environmentally different areas, and (4) develop and conduct technology transfer sessions for the DWR user group.

To monitor the diversified agriculture of California on a statewide basis using Landsat remote sensing techniques requires the definition of

regions where this approach is applicable and where similar techniques may be used. A variety of environments may require a variety or at least a set of remote sensing techniques. The choices considered for this project are Landsat, aerial photography and field work in various combinations and with a variety of temporal permutations. The number of days an area is obscured by cloud cover, the spectral properties of the surrounding native vegetation and the size and shape of fields, especially as affected by surrounding topography, all impact the choice and optimization of the procedural steps. Those areas with the highest target-to-background contrast (either in the spectral, spatial or temporal dimensions) are most amenable to satellite-based remote sensing, while those areas with little contrast presently require a greater dependence on higher resolution aerial photography or field work.

A second criteria for regionalization is both monetary and time costs. For this, areas within the state where the natural environment has historically restricted agricultural development were defined as "minimal crop land zones." This is not to say that remote sensing has no agricultural application in these areas, but generally speaking, the level of agriculture in these areas is so low that the major cost would be the locating of small agricultural areas. On the other hand, where areas of significant size, potential or importance exist, it may prove cost effective to periodically monitor for change detection.

A total of twelve regions of major significance to remote sensing assessment of croplands have been defined. Seven of those regions have areas of significant agricultural importance. The rest, because of climate, topography and lack of good soils, have much less water resource significance and are classified as minimal cropland zones.

In addition to extending the stratification described earlier to the San Joaquin Valley of California and conducting two technology transfer sessions, a number of DWR-defined special interest tasks relating to water use were studied. Although at this point these studies do not relate directly to multiphase estimation of irrigated land, as the APT moves into providing DWR with more detailed water management information studies such as these will be increasingly pertinent. There were four 29,948 hectare (74,000 A) test sites; one located in each of the DWR districts.

In the Northern District, DWR was specifically interested in studying the feasibility of estimating the acreage of land under cultivation for rice early in the growing season. Rice is a high value crop ($167,666,000 in 1977), demands a large amount of water (5-7 acre-feet/year), and the area under production varies considerably, 124,600 - 212,500 hectares (308,000 -525,000 A). If actual rice acreage is less than anticipated, the water which was planned for use early in the season (for initial flooding) and that had been obligated to maintain the surface level throughout the season would then be available for transport and sale elsewhere in the state. Because of this potential water supply, an accurate, early estimation of rice acreage would be very useful to the Department.

Two relatively unique traits governing the location and growth of rice need to be exploited when using Landsat to make an early estimate of

acreage: (1) rice cultivation is confined to areas that are underlain with an impervious subsoil and are generally not suitable for other crops, and (2) the fields are flooded prior to planting and water is visible until the canopy obscures it. The importance of soil type is to enable a stratification of the land into areas where rice production dominates and a "bare soil" signature very early in the season (March to mid-April) can be labeled as rice with some degree of confidence. Initial flooding, which is a more secure indication of rice cultivation takes place in late April through mid-May.

Seven single date and three multitemporal combinations of Landsat color composite imagery were tested. Working with enlarged imagery (1:150,000), 184 dots were randomly located and the analyst was required to label each dot as falling in a rice or non-rice field. To aid in identification, multitemporal full frame Landsat color composite transparencies, the general schedule of rice operations and 1978 adjusted crop calendar were available to the analyst.

For purposes of statistical analysis, the area was divided into four equal size test cells. For each date and date pair, fields interpreted were compared to ground data for omission and commission errors. Both sets of data on percent omission and percent commission error were transformed using the Arcsine Transformation (Sokal and Rohlf 1973) and then were analyzed using a one-way ANOVA test for significant differences between dates and date pairs (Sheffee 1959). When considering omission and commission errors the best date combination for an early estimate was 12 May - 30 May (percent omission error = 1.3%; percent commission error = 2.2%). On this pair, the signature given by the flooded fields was the key to accurate analyst labeling.

For this test site a number of general remarks and recommendations can be made as a result of this analysis: (1) a stratification based on soil type and historical rice cultivation as seen on multitemporal Landsat should significantly reduce the area of estimation and provide a structure for development of an appropriate sampling design, (2) year specific adjustments of the crop calendar are necessary to insure selection of optimum Landsat acquisitions, (3) timely receipt of Landsat data is crucial to making the estimate functionally useful to DWR, and (4) there appears to be every reason to believe that digital analysis of the computer compatible tapes could provide an accurate estimate of rice.

The Central District of DWR has the responsibility to monitor land use and certain field activities in the delta of the Sacramento and San Joaquin Rivers. The combination of rich soil, easily available water, proximity to the San Francisco Bay Area market and convenient shipping led to early development of intensive agriculture. One of the greatest problems facing the Delta farmer is the seasonal build-up of salt in these low-lying soils (approximately 168,000 of the 299,000 hectares (415,000 of the 738,000 A) area lies below sea level).

In order to minimize the effects of salinity or reclaim soils, excess water must be applied to carry salts through the soil and below the root zone. Leaching requires a sufficient quantity of water for salt removal

and the consumptive use of the crop. If leaching takes place during the growing season, 50-100% more water than necessary to meet consumptive use requirements must be applied.

DWR currently monitors the extent of leaching by flying the area in light aircraft, visually identifying the fields being leached and locating on prepared field maps. In the year spanning October 1975 through April 1976, DWR flew and mapped the study area eleven times. Based on the location of cloud cover, six paired data sets (Landsat/DWR map data pairs) were available to study the potential use of Landsat for detecting areas of leaching within the Delta.

Landsat color composite and MSS Band 7 imagery were studied for each of the acquisitions selected. Since standing water normally is quite obvious on Landsat imagery, it had been hoped that the fields being leached would be apparent. Careful study of the imagery on all the aquisitions yielded negative results on the analyst's ability to detect and identify areas of leaching. There are a number of reasons that may have contributed to identification problems encountered in the Delta: (1) the peat soils of this area are normally very dark, (2) the high water table as well as seepage and drainage problems in this area act to keep soils consistently moist, (3) the leaching cycle coincides with the winter rain season and resulting overall soil wetness, and (4) since leaching takes place in the winter, there is a relatively large proportion of land lying fallow. Leaching is the perogative of the individual land owner and a bare soil signature is not a reliable indication of past or potential leaching activities.

The San Joaquin and Southern District sites were both studied for specific crop type determination. The San Joaquin site is dominated by cotton (45%), vineyards (20%) and grain and hay crops (8%). Field crops and orchards make up the remainder. Three dates of imagery, 16 March 1978, 20 July 1978 and 30 September 1978, were selected for analysis. Results of the evaluation indicated that: (1) deciduous orchards and vineyards could not be consistently identified, (2) native vegetation and escaped cultivars represent a confusion class with small grains due to similar phenologies and site degradation of some crop fields by surface drainage and soil salinity, (3) all the major field crops (cotton, small grain, hay and pasture) can be distinguished easily with proper date selection and (4) the seasonality of the cropping patterns, the large size (average 77 ha, 190 A) and regularity of field boundaries favorably impact the use of remote sensing techniques.

The Southern District site was located at the mouth of the Santa Clara River on the Oxnard Plain. Due to a combination of rich soil and mild coastal climate, the area produces a wide variety of crops throughout the year--principally citrus, truck crops and avocadoes. Fields remain fallow for only short periods with new crops generally being planted shortly after harvest. Year-around multicropping results in 2-3 harvests for many fields of truck crops. The most significant acreage was in tomatoes, lemons, strawberries, dry beans, flowers and nurseries, celery, cabbage, bell peppers and lettuce.

While selecting a variety of dates throughout the year was the primary criteria for Landsat image collection, the presence of coastal fog limited the options. Four dates in 1978 were ordered: 24 March, 8 May, 19 July and 29 September. Because of heavy fog on the midsummer date an image form the previous year had to be substituted.

Examination of the imagery revealed numerous interpretation problems arising out of crop characteristics: (1) approximately 40% of the site is used for growing vegetables, (2) nearly 30 different vegetable crops are grown in this area, (3) 80% of the vegetable and truck crop fields are multicropped with both double and triple cropping occurring, and (4) field sizes were generally small (average 4 ha, 10 A) and irregular in shape. Some general conclusions based on the examination are: (1) specific vegetable crop identification will be difficult both because of the number of crops and high level of multicropping, (2) vegetable class identification may be possible with proper date selection, (3) lemon orchards are easily identified, (4) strawberries, because of their stable spatial distribution and growth throughout most of the year can be identified and (5) remote sensing can serve as a technique to monitor urban encroachment into croplands.

2.4 SUMMARY

By the end of 1978, a number of issues critical to DWR had been studied. Of primary concern was the development of a technique by which DWR could produce a statewide estimate of irrigated land using manual analysis of Landsat imagery. Based on earlier work in the ten county study area and the Sacramento Valley test site, a basic methodology for producing the estimate has been formulated. In general, the recommended procedure would include:

- Manual analysis
- Regionalization
- Multitemporal Landsat
  - May
  - July/August
  - September/October
- Multiphase sampling
  - Stratification
  - Regression

Secondary to the manual analysis of irrigated land, preliminary work on the use of digital analysis for estimating and mapping irrigated land was begun on a limited test site basis. Demonstrations of the use of manually interpreted multitemporal Landsat for estimating and mapping specific crop types had also been completed by the end of 1978 (small grains in the Sacramento Valley and Kern County; a variety of field and orchard crops in Kern, Tulare and Ventura counties; and rice in Colusa County).

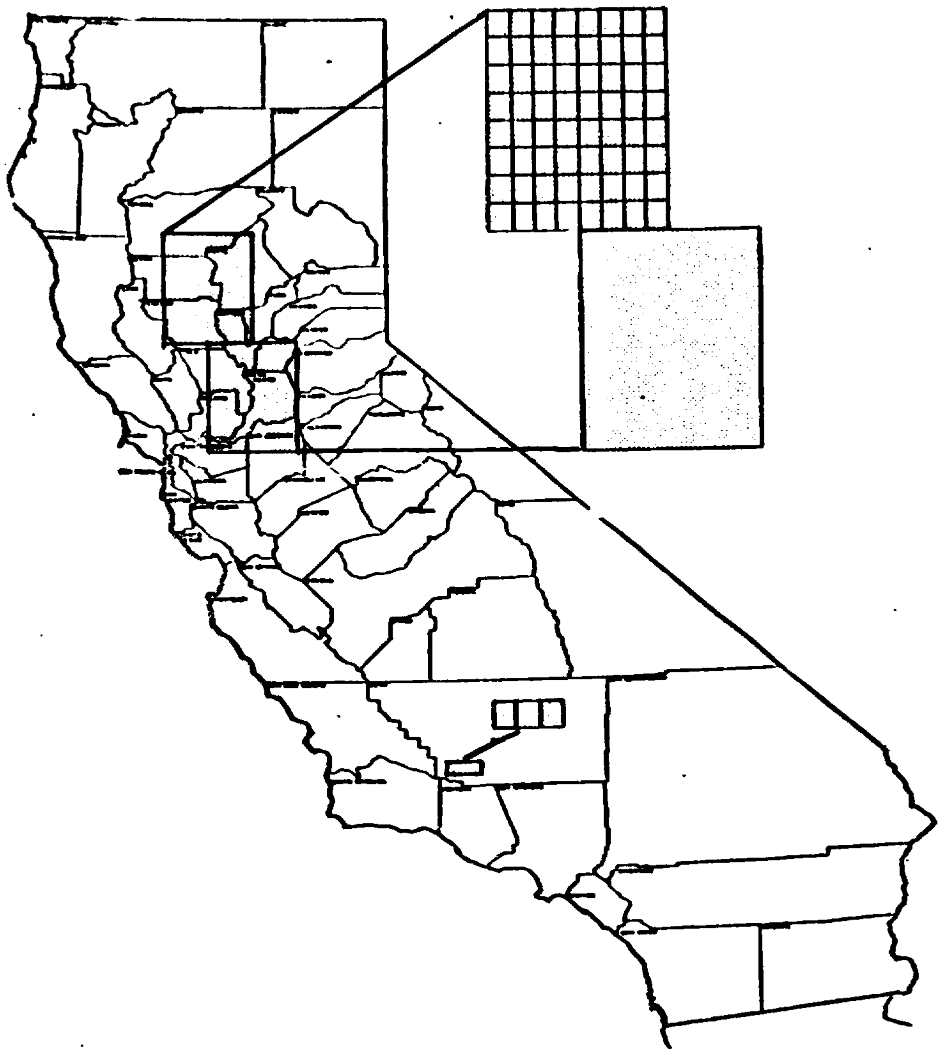
The results of the earlier projects provided the foundation upon which the tasks for 1979 were based. In the following sections the objectives, procedures and results of the work done in 1979 will be discussed. In general, work was divided into four major categories: (1) manual analysis of irrigated lands, (2) digital analysis of irrigated lands, (3) crop type analysis and (4) supporting sampling design. The remainder of the report will address each of these topics in detail.

3.0 1979 - OBJECTIVES

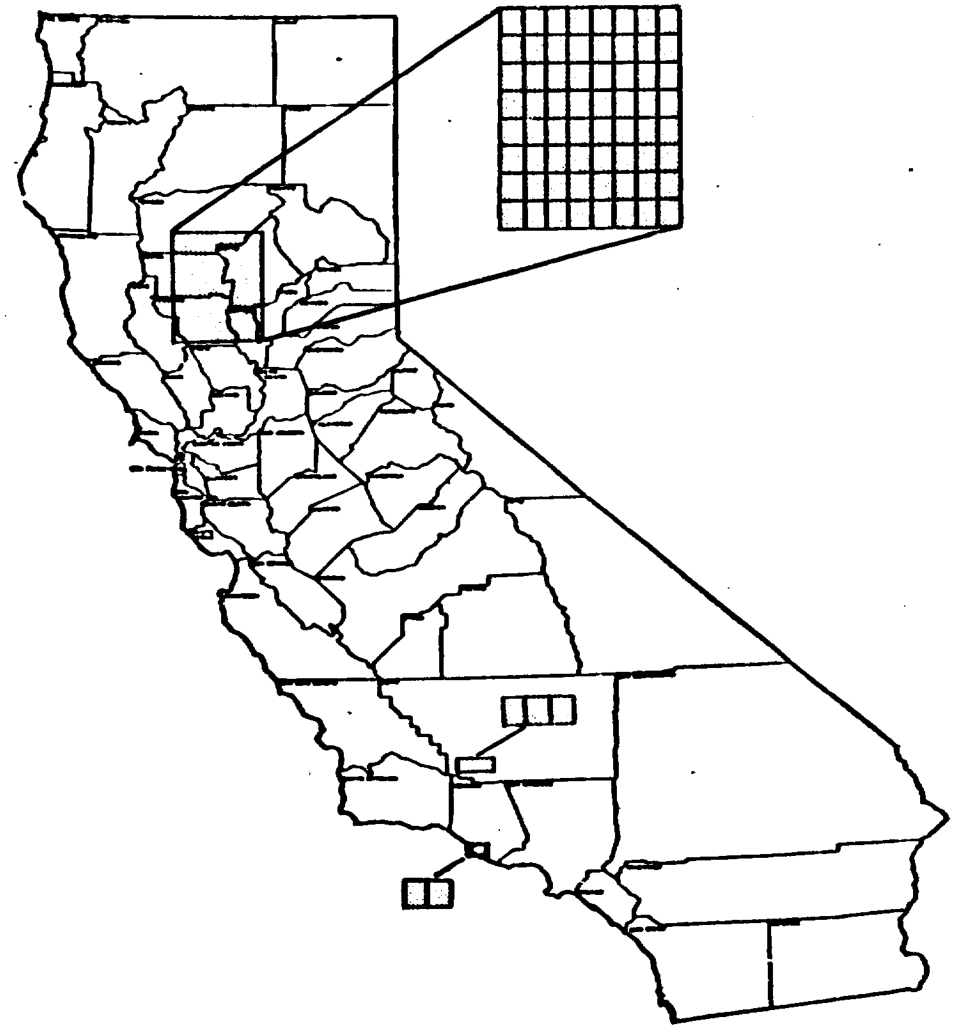
Towards the end of 1978, a number of meetings were held between the Department of Water Resources, NASA/Ames and the University of California to establish task goals and test sites for 1979. Through this cooperative effort, a basic structure incorporating four Landsat data analysis tasks and one sampling design task was constructed:

- Specification of sampling design
- Task I - Estimation of irrigated land using manual analysis techniques
- Task II - Estimation/mapping of irrigated land using digital analysis techniques
- Task III - Estimation/mapping of crop type using manual analysis techniques
- Task IV - Estimation/mapping of crop type using digital analysis techniques.

Using 1979 Landsat data, the principal objective of Task I was to estimate the total irrigated area of the state of California. Using the basic manual analysis methodology developed during the previous projects, this task was designed to test the operational feasibility of producing an accurate estimate of irrigated land over a large area (~40,470,000 ha [~100,000,000 A]), in one year's time (exclusive of planning) and at a reasonable cost. This task dominated the 1979 effort, requiring approximately 40% of the available resources in addition to a majority of the sampling design output. Task II (~20% of 1979 effort) had two major study topics: (1) to investigate potential procedures and associated accuracies with the registration of multitemporal digital Landsat data, and (2) to test various classifications procedures for digital estimation and mapping of irrigated land. Task II used two major test sites, two 1^{o} blocks in the Sacramento Valley and three 7.5' quadrangles in Kern County (see Figure 3-1). Two additional tasks were designed to study crop type identification and mapping using manual (Task III) and digital (Task IV) analysis. In practical operation, Tasks III and IV were treated together (~20% of 1979 work) with greater emphasis put on the digital analysis. The test sites used for crop type work were a 1^{o} block in the Sacramento Valley, three 7.5' quadrangles in Kern County (San Joaquin Valley) and two 7.5' quadrangles in Ventura County (south coast) (see Figure 3-1 for location of the test sites). The final task for 1979 was created to outline sampling design questions for all the tasks and specify a system to be used for the 1979 statewide estimation demonstration (Task I). The Task I work dominated the sampling design effort during the course of 1979, although increasing attention to the other tasks occurred at the end of the calendar year. In this area, as in the data analysis phases, increasing effort will be put on the digital analysis of irrigated land and crop type work as the project progresses in the coming year.



TASK II TEST SITES



TASKS III AND IV TEST SITES

Figure 3-1. Test sites used in the digital analysis of mulitemporal Landsat for the estimation/mapping of irrigated land (Task II) and in the study of the use of Landsat for the estimation/mapping of crop type (Tasks III and IV). Task I work estimated the irrigated land over the entire state of California.

4.0 ESTIMATION OF IRRIGATED LAND USING MANUAL ANALYSIS TECHNIQUES (TASK I)

Successfully producing a highly accurate, repeatable estimate of irrigated land over a state as large as California requires the integration of a variety of components. Based on the experience gained in the two previous projects a set of five sub-tasks was defined to guide the processing of the data from the initial definition of information requirements to production of the estimate. In Figure 4-1, the analysis sub-tasks were organized as follows:

- Design and sample allocation
- Stratification and sample frame construction
- Landsat measurement
- Medium scale photography and ground measurement
- Estimate summary, evaluation and report

For presentational simplicity the remainder of the Task I description will generally follow the five major sub-tasks shown on the analysis flow.

4.1 DESIGN AND SAMPLE ALLOCATION

Specifying the inventory design required addressing several key issues: (1) defining the information required by the California Department of Water Recourses; (2) generating a data set to be used as a preliminary population model to test and refine the previously used estimation system; (3) applying statistical techniques (Monte Carlo) to the data set to simulate model performance; with the simulation testing various mathematical models, evaluating the stratification scheme and determining expected sample sizes for hydrologic basins; (4) specifying the mathematical model, stratification procedures and sample frame for the 1979 inventory; and, (5) computing the actual sample allocation.

4.1.1 Definition of Information Requirements

A necessity in any project is to strictly and accurately define information requirements. This procedure demands frank appraisal by the user agency as to what is really needed and a straightforward explanation of what can be expected from a particular remote sensing system. Certain fundamental questions designed to carefully define DWR's information needs were posed. These questions, and the responses provided by DWR, formed the base upon which the Task I design was built. Table 4-1 briefly summarizes those questions and responses.

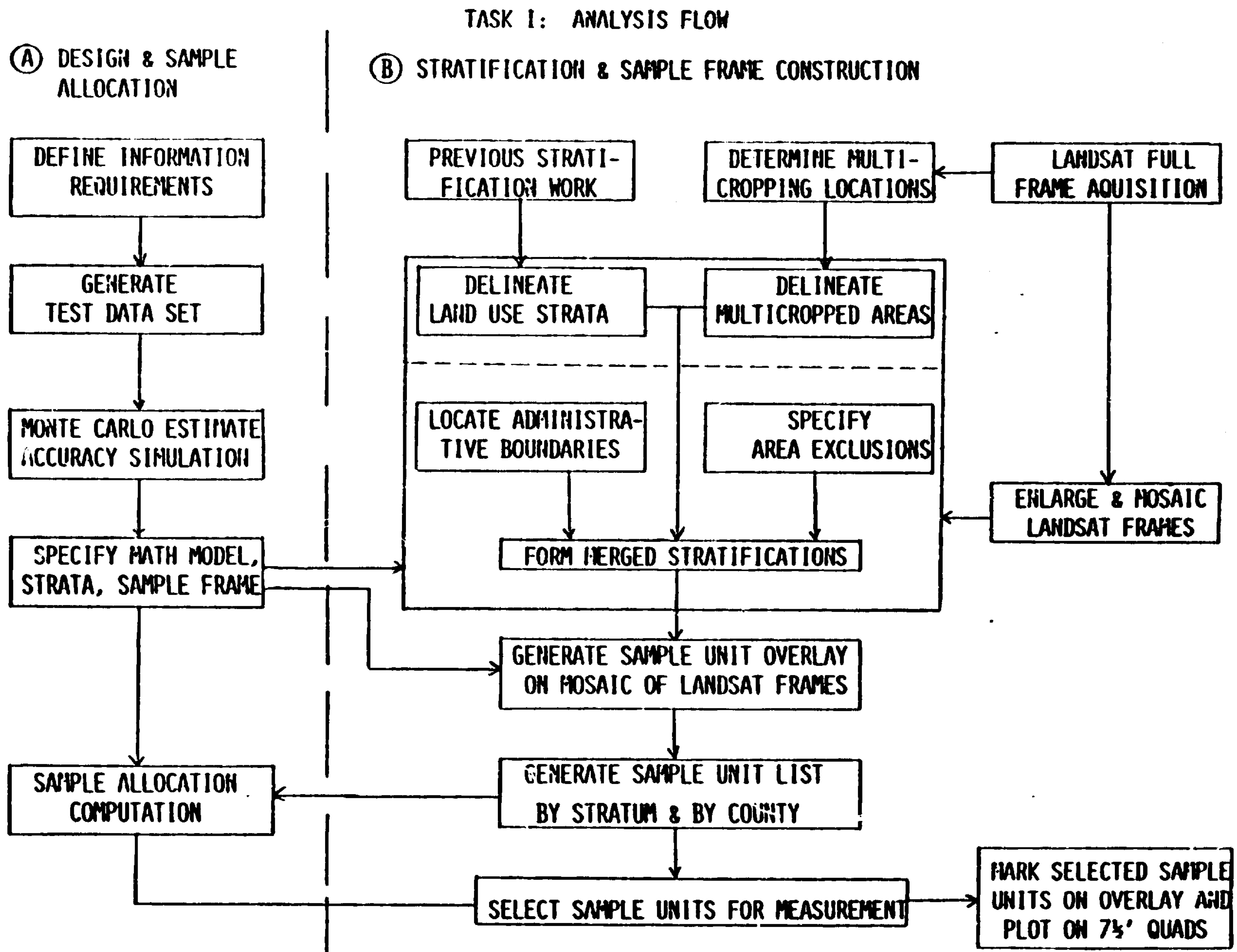


Figure 4-1. Task I analysis flow. Task I was divided into five major organizational sub-tasks: (A)design and sample allocation; (B)stratification and sample frame construction; (C)Landsat measurement; (D)medium scale photography and ground measurement; and (E) estimate summary, evaluation and report.

Ⓒ LANDSAT MEASUREMENT

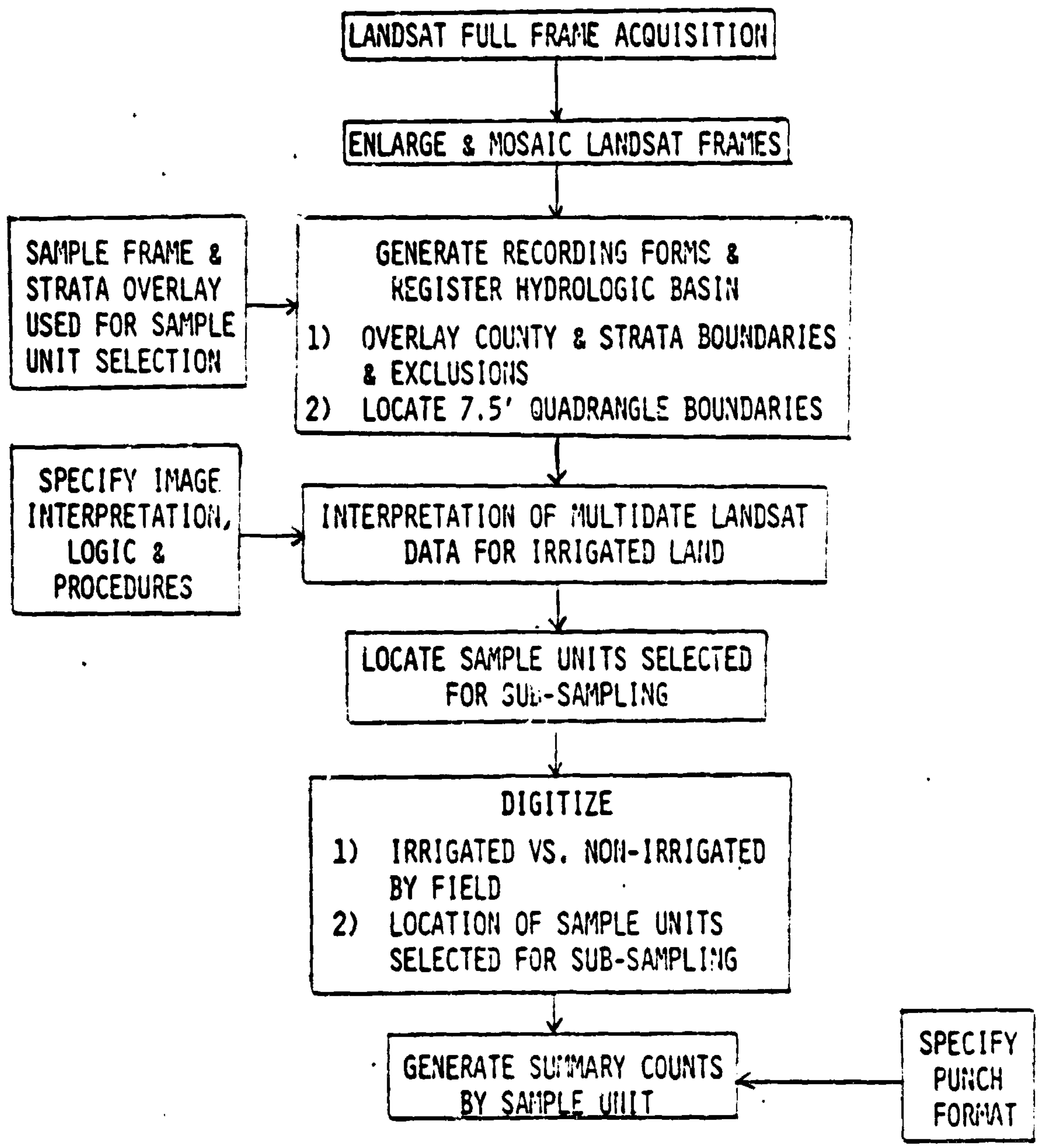


Figure 4-1 (cont'd)

Ⓓ MEDIUM SCALE PHOTOGRAPHY AND GROUND MEASUREMENT

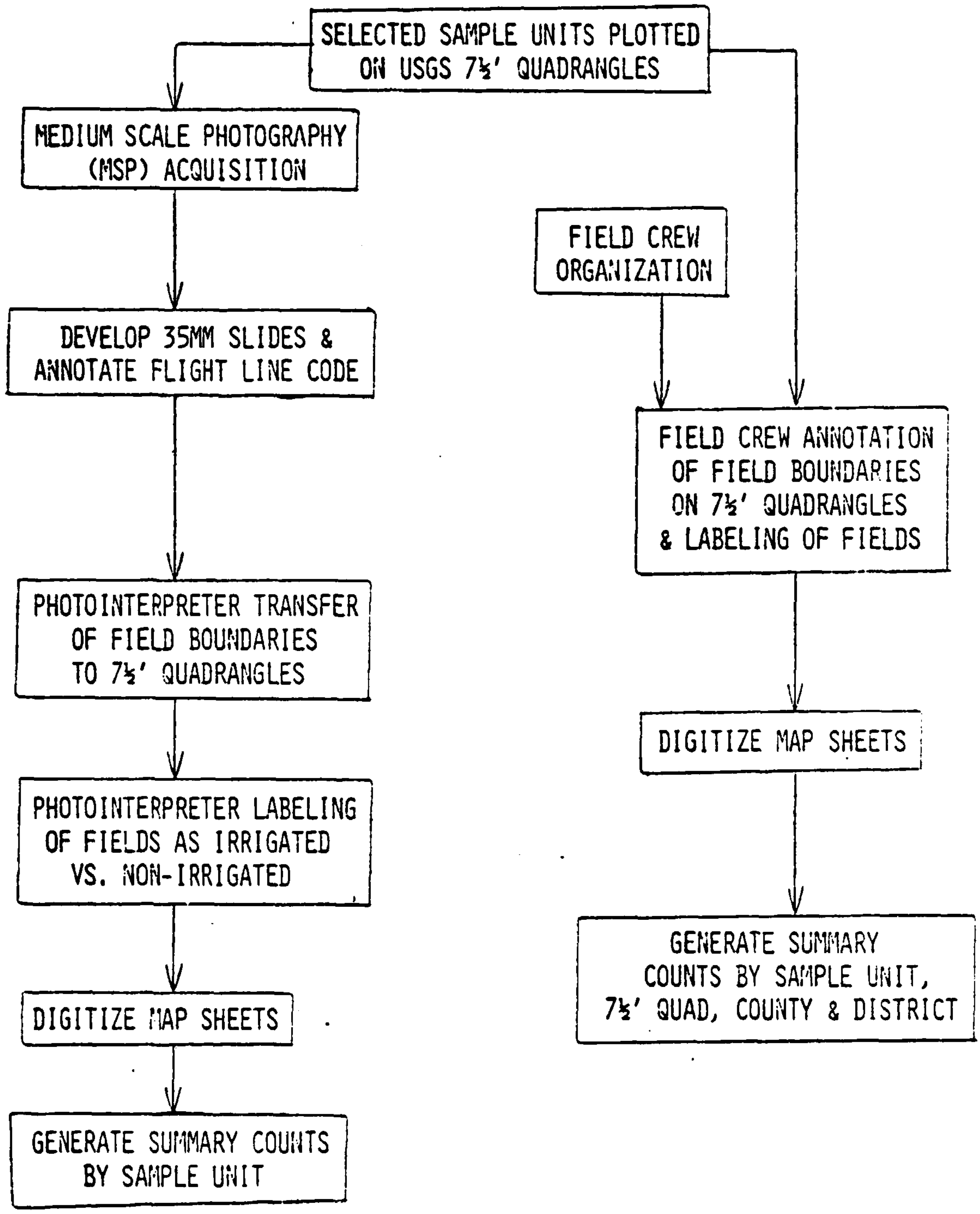


Figure 4-1 (cont'd)

TASK I: ANALYSIS FLOW

Ⓔ ESTIMATE SUMMARY, EVALUATION, REPORT

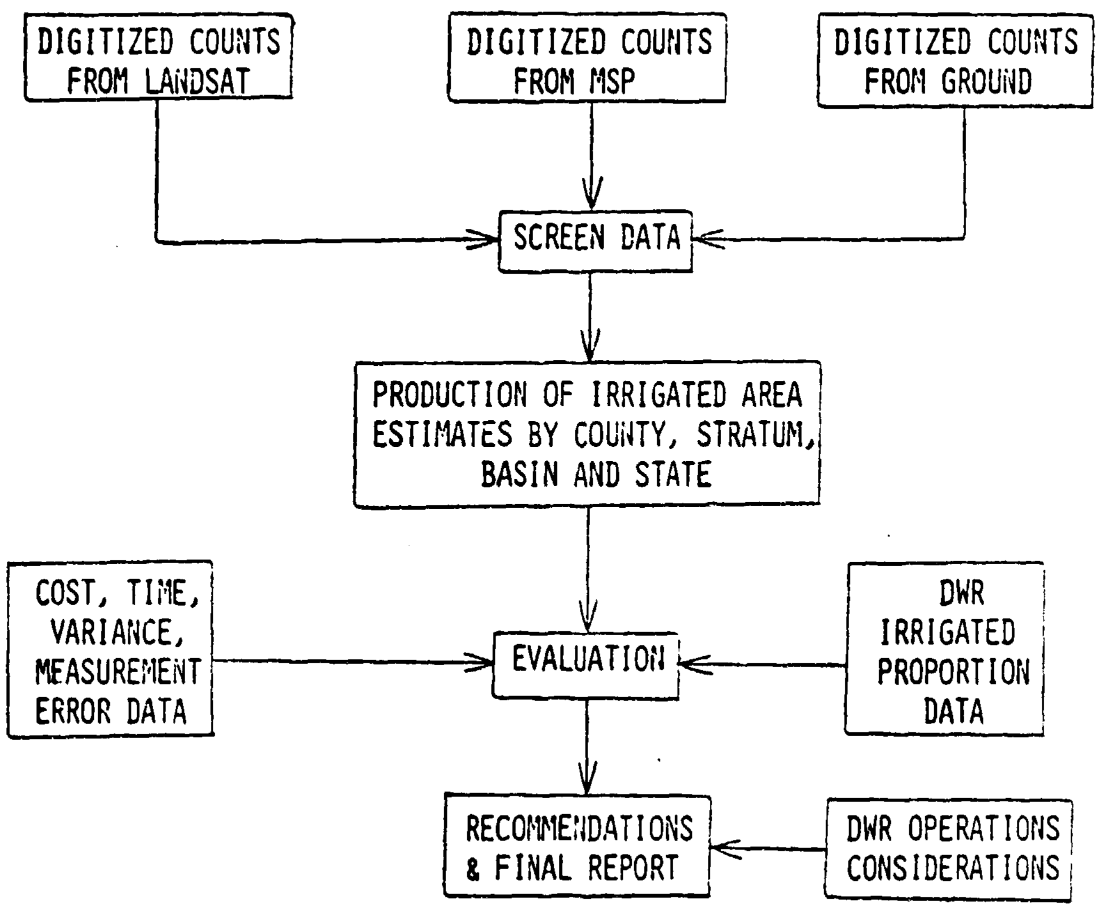


Figure 4-1 (cont'd)

Table 4-1. Design requirements for the statewide estimation of irrigated land.

. Type of information?	. Estimation of the proportion of irrigated land
. Areas of summary?	. Hydrologic Basin (10) . County (58) . State
. Time?	. Inventory data summary within one year, exclusive of planning phase
. Accuracy?	. Estimate precision control at hydrologic basin level . True value of proportion irrigated to fall within ± 5% of estimate 95 times out of 100
. Cost?	. Not formally specified, but in the range of 1 to 2 cents per agricultural acre
. Technology constraints?	. Must be implementable by current DWR personnel and processing capabilities

4.1.2 Generation of Test Data Set

Once inventory information needs were established and understood by all project participants, the sample design phase progressed to the next logical step; evaluating the previously used estimation procedures and developing an improved system to meet the refined and updated inventory objectives of DWR. This evaluation process addressed three major areas of the previously used systems:

- the form and performance of alternative sample system estimators
- the effect of stratification on sampling error, and
- the preliminary computation of sample size for planning purposes

Statistical data collected and analyzed for the 14-County Study was used to address these issues. As described in Section 2.2, 1830 Phase I (Landsat) and 141 Phase II (grd/pi) sample units were selected, allocated by county and and interpreted for estimating the proportion of land irrigated in the 14-County Study area. As the sampling system was a multiphase design, the 141 Phase II units were a locationally-matched subset of the 1830 Phase I units. Table 4-2 shows the proportion irrigated measurements for the matched pairs of sample units. (A complete description of the fourteen county study from which this was derived is given by Wall, Tinney et al, 1978). Using this paired data, Monte Carlo simulations were performed to address the three major points listed above. A detailed description of the methodology and results of the Monte Carlo tests are given in the following section.

4.1.3 Monte Carlo Simulation

Using the population of locationally matched pairs of Landsat and ground data described above, a stochastic technique was used to re-examine three facets of the previous multiphase sampling designs by: (1) testing an alternative mathematical estimator to the regression type used to link estimates made at the various phases; (2) examining the value of the stratification (designed to control measurement error) for controlling sampling error; and (3) computing the approximate number of sample units needed to achieve a percent standard error of $\pm$ 5% at the 95% level of confidence within a hydrologic basin.

Testing Model Alternatives

The statistical technique used to test various alternatives to the regression estimator linking the Landsat interpretation and ground data phases was a Monte Carlo simulation. Being stochastic, Monte Carlo is a process whereby a random sequence of observations (or samples) can be drawn from a population in a repeated fashion. By drawing a large number of samples of variable size from a target population, useful statistics and distributions of that population can be evaluated under various sampling scenarios.

For the current study, the Monte Carlo simulation was used to test the relative performance of two estimators: regression and biased ratio. The biased ratio was evaluated as an alternative since this estimator exhibits lower variance under certain conditions. The form of these estimators, including their variance estimators, is given in Table 4-3.

Table 4-2. Locationally matched pairs of Landsat and ground data measurements of proportion of land irrigated used to compute the estimate of irrigated acreage. This data was used in the 1979 design phase for evaluating alternative estimators, studying the effects of stratification and estimating sample size.

COUNTY	SAMPLE UNIT	PHASE I (LANDSAT)	PHASE II (GROUND DATA)
ALAMEDA	AL4-1	.972	.904
	AL4-2	.828	.841
BUTTE	BU1-1	.372	.355
	BU1-2	.281	.327
	BU2-1	.139	0.000
	BU2-2	.063	0.000
	BU3-1	.959	.383
	BU3-2	.263	.331
	BU4-1	.590	.908
	BU4-2	.707	.744
	BU5-1	1.000	1.000
	BU5-2	.687	.636
COLUSA	CO1-1	0.000	0.010
	CO1-2	.606	.626
	CO2-1	.368	.241
	CO2-2	.368	.160
	CO3-1	.721	.890
	CO3-2	.143	.589
	CO4-1	.663	.856
	CO4-2	.932	.931
	CO4-3	.821	.810
	CO4-4	.812	.877
	CO5-1	.329	.340
	CO5-2	.713	.529
CONTRA COSTA	CC1-1	0.000	0.000
	CC1-2	0.000	0.000
	CC2-1	.679	.720
	CC2-2	.507	.402
	CC3-1	.750	.767
	CC3-2	.791	.779
	CC4-1	.944	.856
	CC4-2	.829	.688
GLENN	GL1-1	.616	.629
	GL1-2	.524	.563
	GL2-1	.407	.459
	GL2-2	.660	.642
	GL3-1	.629	.812
	GL3-2	.398	.905
	GL4-1	.894	.875
	GL4-2	.741	.823
	GL5-1	.712	.910
	GL5-2	.444	.713
	GL6-1	.708	.927
	GL6-2	.776	.860
PLACER	PL1-1	.036	.027
	PL1-2	.059	.067
	PL2-1	.040	.036
	PL2-2	.120	.191
	PL3-1	.215	.173
	PL3-2	.200	.190
	PL4-1	.744	.742
	PL4-2	.621	.722
SACRAMENTO	SA1-1	.041	.112
	SA1-2	.045	[illegible]
	SA2-1	.085	[illegible]
	SA2-2	.196	.222
	SA2-3	.286	.220
	SA2-4	.578	.627
	SA2-5	.543	.525
	SA3-1	.121	.218
	SA3-2	.335	.807
	SA3-3	.486	.544
	SA3-4	.136	.194
	SA3-5	.908	.944
	SA4-1	.894	.940
	SA4-2	.868	.966
SAN JOAQUIN	SA1-1	.543	.552
	SJ1-2	.202	.163
	SJ2-1	.713	.733
	SJ2-2	.970	.942
	SJ3-1	.894	.997
	SJ3-2	.353	.878
	SJ3-3	.836	.917
	SJ3-4	.356	.964
	SJ3-5	.848	.835
	SJ3-6	.932	.671
	SJ3-7	.773	.892
	SJ4-1	.539	.620
	SJ4-2	.935	.973
	SJ4-3	.825	.990
	SJ5-1	.971	.965
	SJ5-2	.858	.818
	SJ5-3	.736	.808
	SJ5-4	.928	.839
	SJ5-5	.962	.391
	SJ5-6	.562	.954
	SJ6-1	.954	.991
	SJ6-2	.877	.875
SHASTA	SH2-1	.190	.173
	SH2-2	.403	.222
	SH3-2	.527	.406
SOLONO	SO1-1	.031	0.000
	SO1-2	0.000	0.000
	SO3-1	.912	.971
	SO3-2	.874	.883
	SO3-3	.928	.993
	SO4-1	.333	.880
	SO4-2	.619	.820
	SO5-1	.928	.817
	SO5-2	.859	.939
SUTTER	SU2-1	.960	1.000
	SU2-2	.362	.853
	SU3-1	.944	.595
	SU3-2	.404	.609
	SU4-1	.945	.809
	SU4-2	.900	.956
	SU4-3	.604	.623
	SU4-4	.796	.942
TEHAMA	TE1-1	.569	.420
	TE1-2	.022	.034
	TE2-1	.865	.822
	TE2-2	.343	.799
	TE2-3	.377	.328
	TE3-1	.859	.726
	TE3-2	.438	.380
	TE5-3	.651	.524
	TE5-4	.660	.718
YOLO	YO1-1	.015	.028
	YO1-2	.035	.053
	YO2-1	.130	.571
	YO2-3	.575	.839
	YO2-4	.694	.955
	YO3-1	.754	.840
	YO3-2	.951	.939
	YO3-3	.872	.994
	YO3-4	.642	.777
	YO3-5	.976	.970
	YO4-1	.993	.929
	YO4-2	.706	.909
	YO5-1	.655	.581
	YO5-2	.943	.990
YUBA	YU1-1	.294	.972
	YU1-2	.487	.515
	YU3-1	.670	.613
	YU3-2	.634	.615
	YU4-1	1.000	.309
	YU4-2	.885	.379
	YU5-1	.371	.931
	YU5-2	.683	.767

Table 4-3. The Monte Carlo simulation was used to compare Regression and Ratio (biased) estimators. Based on Monte Carlo results, SRS and Ratio (unbiased) estimators were deterministically evaluated as possible alternatives.

Simple Random Sample (SRS) (Cochran 1977:18):

$$\hat{Y} = \bar{y} \tag{1}$$

$$V(\hat{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_y^2 \tag{1a}$$

Regression (Cochran 1977:189): $\hat{Y} = b\bar{X} + (\bar{y} - b\bar{x})$ (2)

$$V(\hat{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right) \sigma_y^2 \left(1 + \frac{1}{n-3}\right) (1 - \rho^2) \tag{2a}$$

Ratio - unbiased (Goodman and Hartly 1958):

$$\hat{Y} = \bar{r}\,\bar{X} + \frac{(N-1)\,n}{N\,(n-1)} (\bar{y} - \bar{r}\,\bar{x}) \tag{3}$$

$$V(\hat{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right) \left(\sigma_y^2 + R^2\sigma_x^2 - 2R\mathrm{cov}_{xy} + \frac{\sigma_x^2\sigma_r^2 + \mathrm{cov}_{xr}^{\;2}}{n-1}\right) \tag{3a}$$

Ratio - biased (Cochran 1977:150):

$$\hat{Y} = (\bar{y}/\bar{x})\,\bar{X} \tag{4}$$

$$V(\hat{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right) \left(\sigma_y^2 + R^2\sigma_x^2 - 2R\mathrm{cov}_{xy}\right) \tag{4a}$$

$$\left(1 + \frac{3C_{xx}}{n} + \frac{6C_{xx}}{n}\left[\frac{\rho^2 C_{yy} + C_{xx} - 2C_{xy}}{C_{yy} + C_{xx} - 2C_{xy}}\right]\right)$$

Table 4-3 (continued)

where:

$\hat{Y}$	=	estimate of true proportion irrigated ($\bar{Y}$)
$V(\hat{Y})$	=	estimate of variance of $\hat{Y}$
N	=	population size
n	=	sample size
$\bar{y}$	=	sample mean proportion irrigated for ground data (y_i)
$\bar{X}$	=	population mean proportion irrigated for Landsat data (x_i)
$\bar{x}$	=	sample mean proportion irrigated for Landsat data (x_i)
$\bar{r}$	=	sample mean for ratio $y_i/x_i = r_i$
R	=	true ratio of $\bar{Y}/\bar{X} = \hat{Y}/\bar{X}$
ρ	=	sample correlation between x_i and y_i
$\hat{\sigma}_x^2$	=	sample variance of proportion irrigated for Landsat data
$\hat{\sigma}_y^2$	=	sample variance of proportion irrigated for ground data
$\hat{\sigma}_r^2$	=	sample variance for ratio $y_i/x_i = r_i$
cov_{xy}	=	sample covariance of x_i and y_i
cov_{xr}	=	sample covariance of x_i and r_i
C_{xy}	=	$cov_{xy}/(\bar{x}\ \bar{y})$
C_{xx}	=	$\hat{\sigma}_x^2/\bar{x}^2$
C_{yy}	=	$\hat{\sigma}_y^2/\bar{y}^2$
b	=	regression coefficient of y on x

Using the population of 141 matched pairs of Phase I and Phase II proportion measurements (Table 4-2), a large number of samples were drawn by stratum for Monte Carlo simulation. In addition to calculating a set of basic statistics by stratum, the performance of both the regression and ratio estimators was tested by computing the average bias and sampling error. Performance was tested at both the stratum- and 14 County-level for various sample sizes. The general description of the simulation is shown in Figure 4-2.

A summary of the results of the Monte Carlo simulation are tabulated in Table 4-4. Three levels of simulation are shown: individual stratum, selectively aggregated strata, all strata combined. For each stratum and combined strata, the average bias and sampling error were tabulated with confidence limits specified for each level simulated. The average bias and sampling error were calculated using the formulas:

$$\text{average bias} = \frac{\Sigma\,(\hat{Y} - Y)}{m} \tag{5}$$

$$\text{sampling error} = \sqrt{\frac{\Sigma(\hat{Y} - Y)^2}{m - 1}} \tag{6}$$

where, $\hat{Y}$ = estimate of true proportion irrigated ($\bar{Y}$) from either the ratio or regression estimator

Y = ground truth based on DWR-collected data during the 14-County Study

m = number of Monte Carlo iterations (50)

As the sensitivity of both the ratio and regression estimators was based on the expected relative bias and sampling error, two important observations can be made by reviewing the tabulated results. The first observation is that stratification appeared not to have significantly reduced sampling error. This observation will be addressed in more detail below. The second observation is the apparent lack of any significant difference between the two estimators' performance as exhibited by the very similar values of average bias and standard errors of the estimate at both the 95% and 99% levels of confidence. Except for small sample sizes, the regression estimator exhibited lower bias and variance than did the ratio estimator. Though the regression estimator was judged superior to the ratio over most strata, the results of this Monte Carlo did not clearly indicate which estimator, if either, was the better to accomplish the objectives of the state-wide inventory of irrigated land proportion. Based on these Monte Carlo results, a more in-depth analysis of other mathematical estimators was performed.

Using the same data set from the 14-County Study, two additional estimators were evaluated with the ratio (biased) and regression forms. These two estimators, the simple random sample (SRS) and Ratio (unbiased), are shown in Table 4-3. The performance of all four estimators was evaluated deterministically by predicting ground variance (σ_y^2) and correlation between Landsat and ground data (ρ^2).

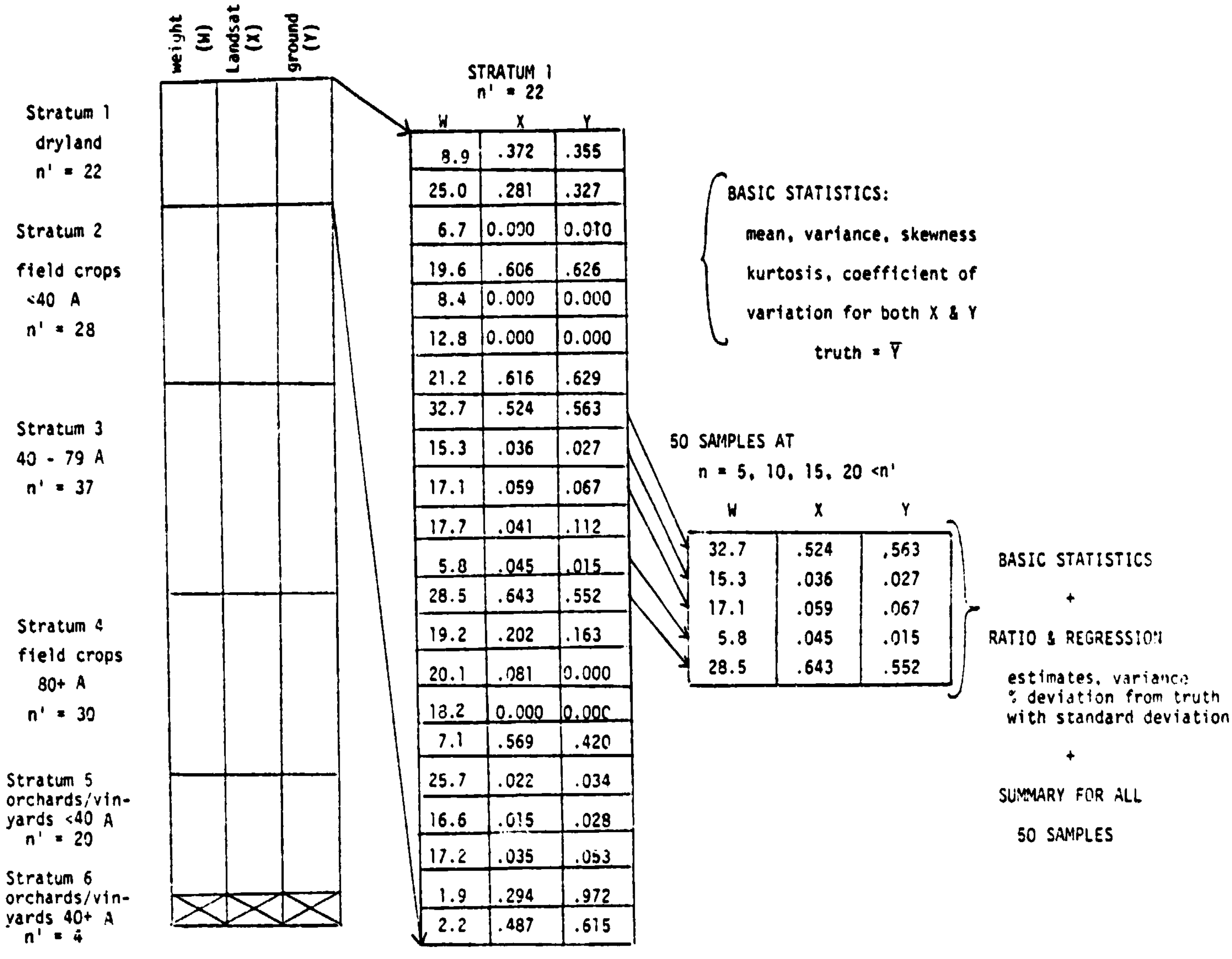


Figure 4-2. General description of the Monte Carlo simulation. The 141 matched pairs of Landsat and ground data from the 14-County study were broken into their six strata. From each stratum (stratum 1 shown as an example) samples were repetitively drawn (50 times) to generate estimates for various sample sizes: 5, 10, 15, 20.... (one sample of size 5 is shown as an example).

Table 4-4. Summary of selected results of the Monte Carlo Simulation used for Task I.

Stratum	Proportion Irrigated from Ground Data	Sample Size	Population Size (Weight)	Proportion (Relative Weight)	Estimators: Ratio Bias ± SD	Estimators: Regression Bias ± SD
1	.249	15	227	.1240	.0063 ±.0611	-.0031 ±.0372
2	.504	20	221	.1208	.0045 ±.0381	.0023 ±.0375
3	.736	15	664	.3628	.0014 ±.0330	.0053 ±.0270
4	.856	10	516	.2820	.0065 ±.0371	.0071 ±.0264
5	.830	10	137	.0749	.0026 ±.0039	.0035 ±.0350
6	.830	10	65	.0355	.0026 ±.0039	.0035 ±.0350
Average	.692	80	1830	1.000	.0039 ±.0184	.0042 ±.0142
Confidence Limits					±5.3 @ 95 ±7.0 @ 99	±4.1 @ 95 ±5.4 @ 99
1	.249	15	227	.1240	.0063 ±.0611	-.0031 ±.0372
2	.504	20	221	.1208	.0045 ±.0381	.0023 ±.0375
3-4	.788	30	1180	.6448	.0025 ±.0188	.0060 ±.0190
5-6	.830	15	202	.1104	-.0011 ±.0213	-.0011 ±.0217
Average	.692	80	1830	1.000	.0028 ±.0152	.0037 ±.0141
Confidence Limits					±4.4 @ 95 ±5.8 @ 99	±4.1 @ 95 ±5.4 @ 99
All Strata Combined (1-6)	.692	80	1830	1.000	.0021 ±.0129	.0029 ±.0125
Confidence Limits					±3.7 @ 95 ±4.9 @ 99	±3.6 @ 95 ±4.[illegible] @ 99

By determining the variance of the estimators using variable sample sizes, the relative performance of each estimator could be evaluated. That estimator exhibiting the lower variance for given sample sizes would be preferred for the state-wide inventory. These estimators were chosen for performance evaluation after an extensive review of the statistical literature. Under certain conditions, the estimator for SRS and/or ratio (biased) estimation could achieve a smaller variance for a given sample size (Wensel, 1977). An alternative ratio estimator, described by Goodman and Hartly (1958), was selected because it is unbiased, and provides a good comparison with the previously evaluated biased ratio estimator. The performance of the estimators is graphically displayed in Figure 4-3a-d. Three of the estimators (SRS, regression, and unbiased ratio) were compared to the biased ratio having an assumed standardized variance of 1.0. Variances were calculated for each of four strata (1, 2, 4, 5-6) with sample sizes ranging from 2 to 25, including an estimation of variance for very large sample sizes ($n \to \infty$). The variances were standardized by dividing by the variance of the biased ratio estimator for the corresponding stratum and sample size.

By examining variance plotted against sample size (Figures 4-3 a-d), it can be seen that the regression estimator was superior to all other estimators for large sample sizes ($n \geq 5$). The SRS estimator was consistently inferior; at all sample sizes SRS standardized variances greater than 3.0 were frequently observed but not plotted on the referenced figures. For small sample sizes both ratio estimators are superior to regression but indistinguishable from each other. Because the standard error of the biased ratio estimator was at most 13% less than that of the unbiased ratio estimator, and given the advantages of using an estimator with no bias, the unbiased ratio estimator was used for small sample sizes. For operational application of these two estimators in a variable sample size environment, specific decision rules must be established for sample size computation when using either estimator. Table 4-5 lists the range of sample sizes (n) over which the unbiased ratio and regression estimators are used.

Table 4-5. Range of sample sizes (n) over which the unbiased ratio and regression estimators are used.

Stratum	Unbiased Ratio	Regression
1	$n \leq 3$	$n \geq 4$
2	$n \leq 10$	$n \geq 11$
3	$n \leq 5$	$n \geq 6$
4	$n \leq 5$	$n \geq 6$
5	$n \leq 6$	$n \geq 7$
6	$n \leq 6$	$n \geq 7$
7	$n \leq 5$	$n \geq 6$

Note: Strata 3 and 7 were not used in the 1976 study but were part of the current inventory. Too few observations existed for Stratum 6; units in this stratum were combined with Stratum 5 for computational purposes. The values given for Strata 3, 6, and 7 are based on the results of those strata with the most similar characteristics.

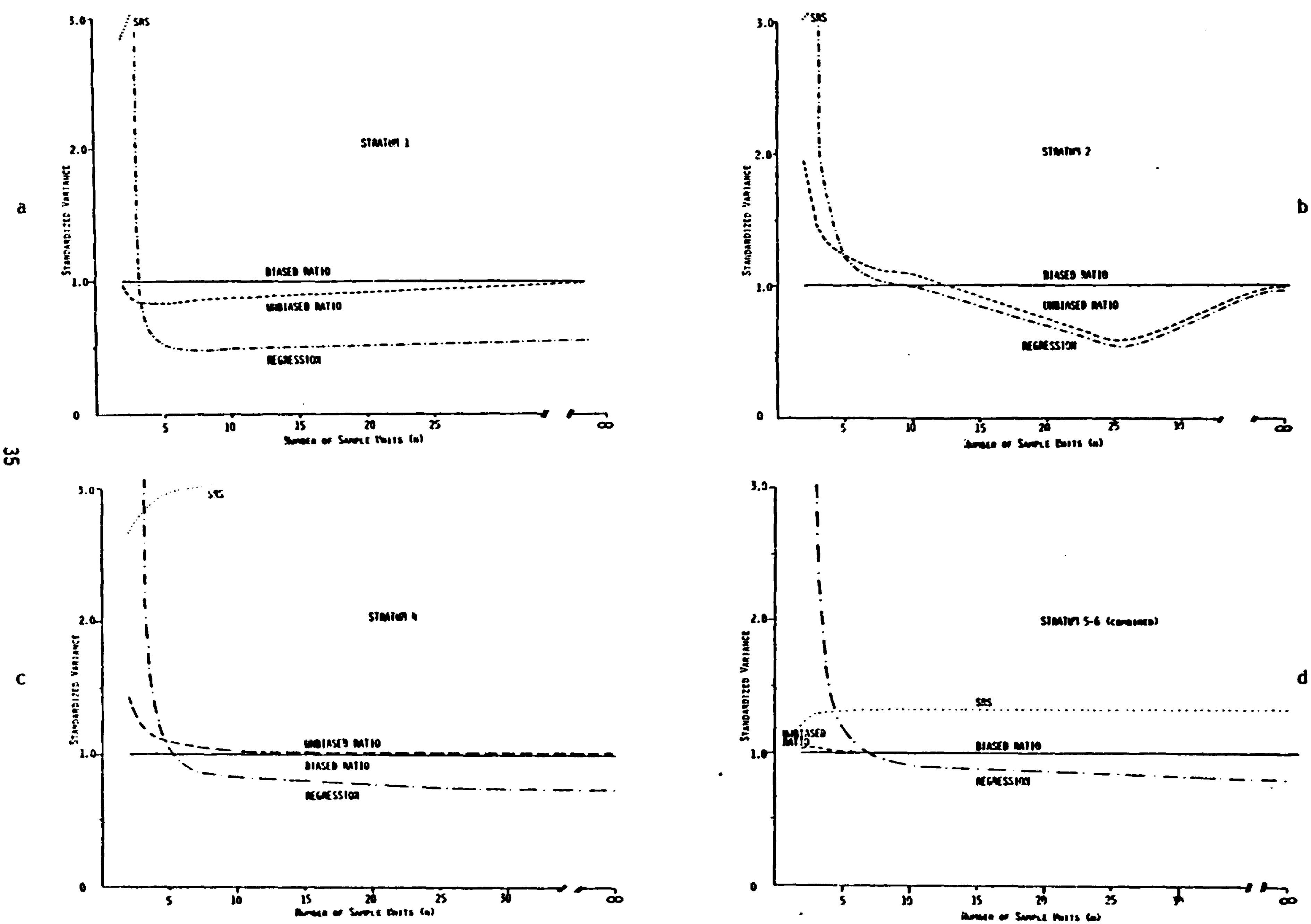


Figure 4-3. Standardized variance as a function of sample size for the SRS, ratio (biased and unbiased) and regression estimators. This data was used to select the mathematical model used to link the phases and to define the sample sizes over which they would be used.

The use of a ratio estimator does present one technical problem: the calculation of $r_i = y_i/x_i$ when either or both x_i and y_i are zero. In such cases the following contingency table is used:

x_i \ y_i	$= 0$	$\neq 0$
$= 0$	$r_i = 1$	$r_i = \frac{1}{1 - y_i}$
$\neq 0$	$r_i = 1 - x_i$	$r_i = y_i/x_i$

In summary, four estimators were evaluated using both stochastic and deterministic tests to ascertain which estimators generated the lowest variance over the range of sample sizes needed for a state-wide inventory system. Both the previously used regression and a newly tested unbiased ratio estimators were considered to be the best estimators as they exhibited the lowest variances. Depending on the stratum sample size, as shown in Table 4-5, the appropriate estimator will be used to calculate the proportion of irrigated land.

Effect of Stratification on Sampling Error

The second objective of the Monte Carlo tests was to evaluate the agricultural practice stratification used in the Sacramento Valley fourteen-county site. Since stratification can potentially reduce variance without increasing total cost (Wensel, 1977), a careful examination of the results of the stratification was warranted. An additional sub-task was also undertaken to study the way in which regression estimators interact with stratification.

A stratification based on land use and field size had been used for the Sacramento Valley estimation (Section 2.2). It was designed to control measurement error associated with the interpretation of agricultural environments that vary considerably in "ease" of interpretation and accurate line placement. The major purpose of the Monte Carlo simulations in this instance was to evaluate the utility of the 14-County stratification in reducing sampling error as well as measurement error.

The regression estimator used in Task I enables a small amount of costly data (i.e., ground survey) to be used in conjunction with a large amount of less costly data (i.e., remotely sensed) in a way that the ground can correct for bias in the remotely sensed; while Landsat can compensate for the small ground sample size, thus reducing sampling variance (Wensel, 1977; Thomas, 1979). Combining such an efficient estimator with a potential error reducing stratification method seemed advantageous. When the stratification and regression estimator were combined in the Monte Carlo simulations, however, stratification did little to reduce sampling variance. This minimal effect on variance can be seen by evaluating the results shown in Table 4-4. Those results indicate that in neither of the two stratification levels simulated (6-strata, and selectively combined strata) was there a significant difference in variance compared to the case where no stratification was used. Since these results would have significant implications for this and future studies, further investigation of the regression estimator variance equation was warranted.

For the unstratified case, V(Y)un, the variance formula is (Refer to Table 4-3 for notation):

$$V(Y)_{un} = (\frac{1}{n} - \frac{1}{N})\ (1 + \frac{1}{n-3})\ (1 - \rho^2)\ \sigma_y^2 \qquad (7)$$

While for the stratified case, $V(Y)_s$, we have:

$$V(Y)_s = \sum_{i=1}^{\lambda} W_i^2\ (\frac{1}{n_i} - \frac{1}{N_i})\ (1 + \frac{1}{n_i-3})\ (1 - \rho_i^2)\ \sigma_{y_i}^2 \qquad (8)$$

Consider the term $(1 - \rho^2)\ \sigma_y^2$ in isolation:

$$(1 - \rho^2)\ \sigma_y^2 = (1 - \frac{RSS}{TSS})\ \sigma_y^2 = (1 - \frac{TSS - ESS}{TSS})\ \sigma_y^2$$

$$= (\frac{ESS}{TSS})\ \sigma_y^2 = (\frac{MSE(n-2)}{\sigma_y^2(n-1)})\ \sigma_y^2$$

$$= MSE\ (\frac{n-2}{n-1})$$

Where,

RSS = regression sum of squares

ESS = error sum of squares

TSS = total sum of squares = RSS + ESS

MSE = mean square error

W_i = proportion of basin composed of stratum i

Substituting back into the original variance equations and making the following assumptions:

1) the strata are all the same size, thus $N_i = N$;
2) the strata all have the same mean square error, thus $MSE_i = MSE$;
3) from 1) and 2), optimal allocation would give $n_i = n$;
4) from 1) it follows that $W_i = 1/N$
5) the total population size for the unstratified case is the product of the number of strata times the strata size = λN;

gives the new unstratified variance:

$$V(Y)_{un} = \left(\frac{1}{\lambda n} - \frac{1}{\lambda N}\right)\left(1 + \frac{1}{\lambda n-3}\right)\left(\frac{\lambda n-2}{\lambda n-1}\right) MSE_{un} \tag{9}$$

and the new stratified variance:

$$V(Y)_{s} = \left(\frac{1}{\lambda n} - \frac{1}{\lambda N}\right)\left(1 + \frac{1}{n-3}\right)\left(\frac{n-2}{n-1}\right) MSE_{s} \tag{10}$$

Comparing the two variance equations shows that for small sample sizes the unstratified variance can be lower because:

$$\left(\frac{\lambda n-2}{\lambda n-1}\right)\left(1 + \frac{1}{\lambda n-3}\right) \leq \left(\frac{n-2}{n-1}\right)\left(1 + \frac{1}{n-3}\right)$$

for large n, both these terms approach unity and the only difference between the two equations is the value of MSE_{un} and MSE_s. Thus, stratification will significantly decrease the variance only if MSE_s is significantly less than MSE_{un}. This will only occur if the stratified regressions are a significant improvement over the single unstratified regression. This may not be the case in either of Tasks I or II. Because the Landsat estimate and the ground estimate for each sample unit tend to be equal, the regressions all tend toward a slope of 1.0 and an intercept of 0.0.

Thus in practice, stratification may not give a significantly better fit, and consequently may not give a significant reduction in variance. Stratification will help if strata can be identified that have different biases. After reviewing the Monte Carlo results, the stratification was redesigned as described in Section 4.1.4 in hope of achieving differing regressions.

Preliminary Sample Unit Computation

The third major function of the Monte Carlo simulation was to compute the approximate number of sample units that would be needed to achieve the stated accuracy requirements (±5% at the 95% confidence level for each hydrologic basin). As DWR was responsible for collection of ground (Phase II) sample data, the preliminary computation of sample size was to provide a guideline for planning DWR manpower requirements. Computation of the final sample size for the operational inventory is discussed in Section 4.1.5.

For each sample size (n=5, 10, 15, 20...) used in the Monte Carlo simulation, the number of samples (n^*) that fell within 5% and 10% of the true estimate was determined. This number was converted to a percentage by dividing by the number of cycles (m) and multiplying by 100 ($\frac{n^*}{m}$ X 100). The results are illustrated in Figures 4-4a to 4-4h. Preliminary sample sizes were predicted from these graphs by: (1) stratum, (2) selectively combined strata, (3) no strata, (4) hydrologic basin, and (5) state. Based on this preliminary analysis a maximum number of 80 units (Table 4-4) per hydrologic basin (800 units for the state) was used by DWR for planning the allocation of manpower.

4.1.4 Specification of the Mathematical Model, Stratification Scheme and Sample Frame

The Monte Carlo simulations described in Section 4.1.3 provided the information needed to refine the mathematical estimators used to link multiphase measurements for producing the Task I estimate of irrigated acreage. The simulations also indicated that modifications to the stratification scheme would be necessary if the stratification was to be used to reduce sampling as well as measurement error. The sampling frame remained similar to that used in the previous studies (cluster sample units, 1.6 x 8.0 kilometers in size [1 x 5 miles]) although work by Arno (1979) and UCSB (Appendix I) offered alternatives for further investigation.

Specification of the Mathematical Model

As in the previous studies, the primary equations (estimators) used to link Landsat and ground area measurements to produce estimates of irrigated area were of the linear regression type. The general form of these equations, as adapted to the irrigated lands problem, was established in the original ten county study (Section 2.1). In that study, a multiphase sampling scheme (Tikkiwal 1955 and 1967) was adapted to using iterative estimators whereby the ground (Phase III) estimator used the aerial photo (Phase II) estimator which, in turn, used the Landsat (Phase I) estimator. In both the present and the 14-county study, only two phases were employed: a census at the Landsat phase (Phase I) and a simple random sample within strata at the ground phase (Phase II).

The estimators are affected by the fact that the sample units are considered as clusters and that these clusters are of unequal size. As the clusters were of unequal size, accurate measures of the sizes of the individual sample units were required so that weighted means could be used in the estimators rather than unweighted means. Therefore, the Phase I estimator is:

$$\hat{Y}^* = \frac{1}{n^*}\sum_{i=1}^{n^*} Y_i^* \left(\frac{M_i}{\bar{M}^*}\right) = \frac{1}{n^*}\sum_{i=1}^{n^*}\frac{a_i^*}{\bar{M}^*} = \frac{\sum_{i=1}^{n^*} a_i^*}{\sum_{i=1}^{n^*} M_i} \tag{11}$$

The Phase II estimator is:

$$\hat{Y}' = \frac{\sum^{n'} a_i'}{\sum^{n'} M_i} + \hat{\rho}_{Y^*,Y'}\left(\frac{\hat{\sigma}_{Y'}}{\hat{\sigma}_{Y^*}}\right)\left(\hat{Y}^* - \frac{\sum^{n'} a_i^*}{\sum^{n'} M_i}\right) \tag{12}$$

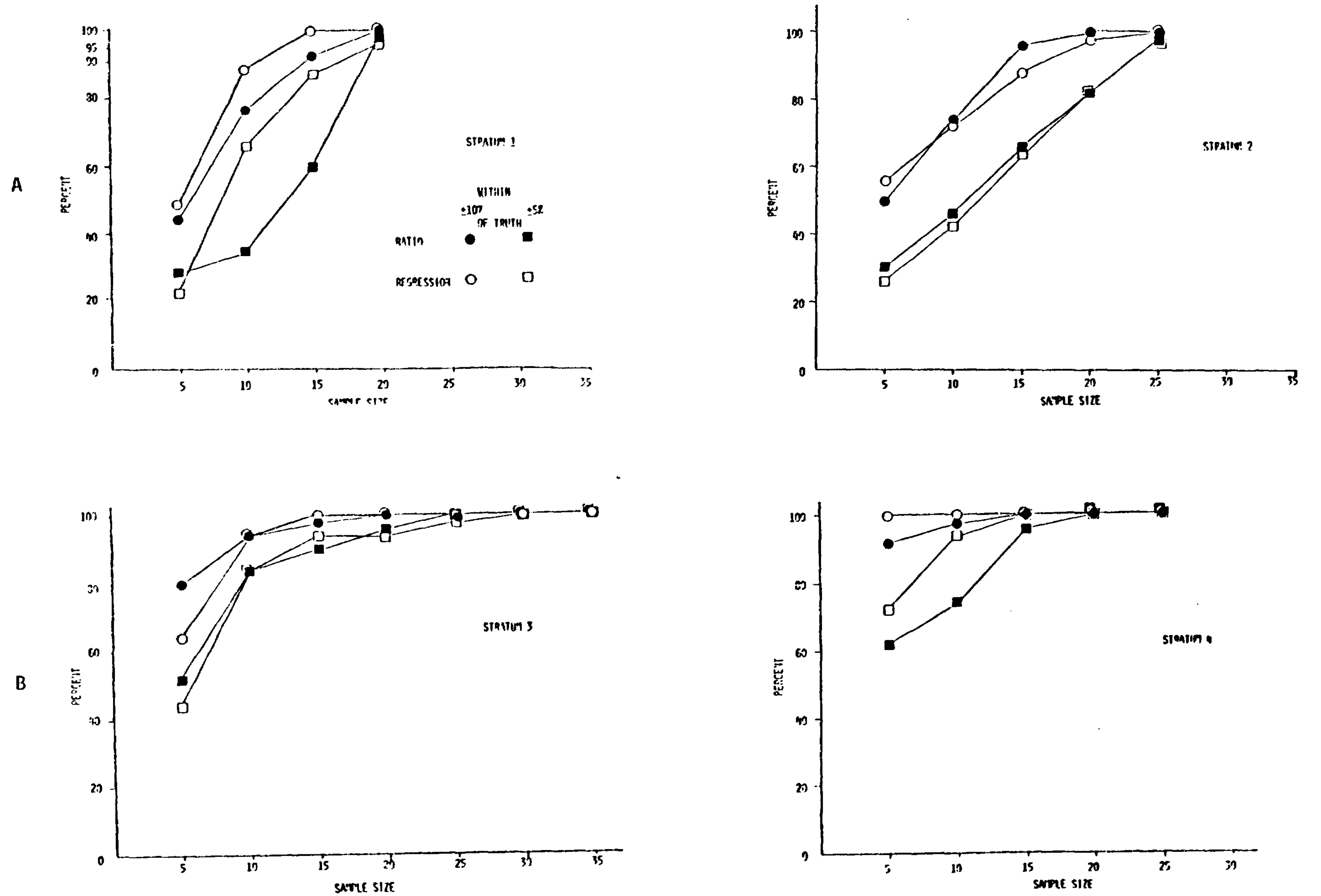


Figure 4-4. The effect of sample size on accuracy of the biased ratio and regression estimators from the Monte Carlo study. Accuracy is measured as the percent of the estimates that fell within ±5 or ±10% of ground truth.

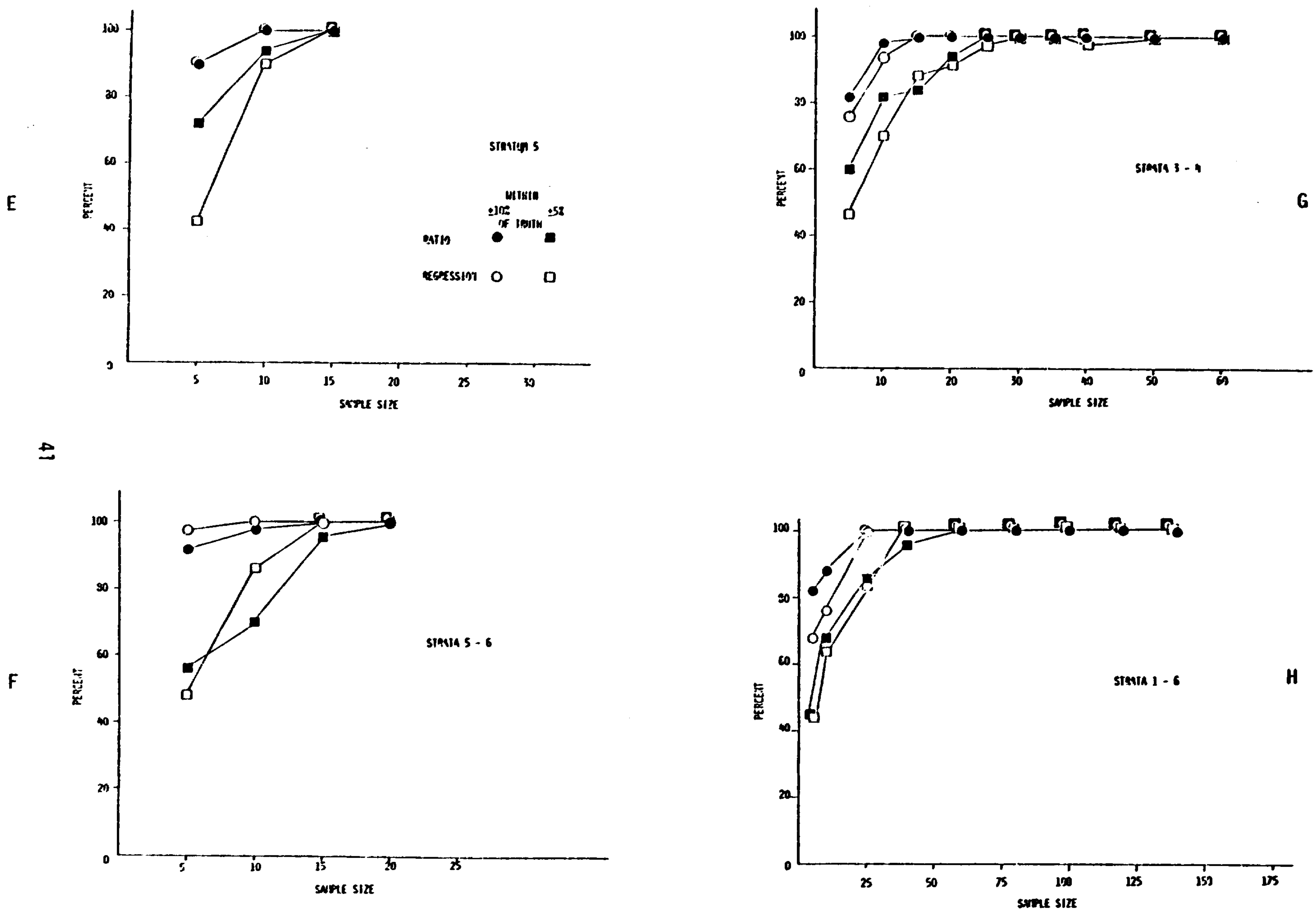


Figure 4-4 (continued)

where:

N	Population size of units to be sampled
n^*	Phase I (LANDSAT) sample size
n'	Phase II (Ground) sample size
M_i	Size of sample unit i (any consistent unit of measure)
$\bar{M}^*$	Mean Phase I sample unit size; $\bar{M}^* = \frac{1}{n^*} \sum^{n^*} M_i$
a_i^*	Irrigated area in sample unit i of Phase I
a_i'	Irrigated area in sample unit i of Phase II
Y_i^*	Irrigation proportion in sample unit i of Phase I; $Y_i^* = a_i^*/M_i$
$\hat{\sigma}_{Y^*}$	Sample standard deviation for weighted Phase I observations
$\hat{\sigma}_{Y'}$	Sample standard deviation for weighted Phase II observations
$\hat{\rho}_{Y^*,Y'}$	Sample correlation between weighted Phases I and II

Note that Equation 12 uses the Phase I estimator $\hat{Y}^*$. The first term is the weighted Phase II mean and the second is its regression correction. The regression coefficient is the term involving the correlation and the standard deviations. It may be seen from this that higher correlations between Phases I and II increase the effect of the correction term (which may be either positive or negative). Also, the smaller the Phase II standard deviation is in relation to the Phase I standard deviation, the smaller the effect of the correction term becomes.

The variance estimates are also computed in an iterative manner. The Phase I estimator is simply the variance of the weighted observations for simple random sampling with a finite population:

$$\hat{VAR}\ (\hat{Y}^*) = \hat{\sigma}^2_{Y^*} \left(\frac{1}{n^*} - \frac{1}{N} \right) \tag{13}$$

The second phase variance estimator is:

$$\hat{VAR}\ (\hat{Y}') = \hat{\sigma}^2_{Y'} \left[\left(\frac{1}{n'} - \frac{1}{n^*} \right) \left(1 + \frac{1}{n' - 3} \right) \left(1 - \hat{\rho}^2_{Y^*,Y'} \right) + \frac{\hat{VAR}\ (\hat{Y}^*)}{\hat{\sigma}^2_{Y^*}} \hat{\rho}^2_{Y^*,Y'} \right] \tag{14}$$

This depends directly on the Phase II standard deviation and uses the Phase I variance estimate. The variance equation (14) differs slightly from the forms used in both the 10- and 14-county studies. The finite population correction factor has been changed to $\left(\frac{1}{n'} - \frac{1}{n^*} \right)$, and a small sample size correction factor has been included: $\left(1 + \frac{1}{n' - 3} \right)$.

In the present study, the Landsat area measurements constituted a census of the sample unit population (i.e. $n^* = N$). Thus, $\hat{VAR}\ (\hat{Y}^*) = 0$ and Equation 14 collapses to:

$$\hat{VAR}\ (\hat{Y}') = \hat{\sigma}^2_{Y'} \left(\frac{1}{n'} - \frac{1}{N} \right) \left(1 + \frac{1}{n' - 3} \right) \left(1 - \hat{\rho}^2_{Y^*,Y'} \right) \tag{15}$$

In order to calculate the irrigation area estimates, a FORTRAN program, MPHASE, had been written previously to compute three phase estimates and the associated variance estimates. In the absence of a third level of information, MPHASE can be used for <u>two phase estimates</u> as well. In either case, there is the option to combine the observations from different strata for the two phases with the least observations in order to obtain more stable standard deviation and correlation estimates. The program was designed to use as many as seven variables of interest per run, so that variables other than irrigated proportion (i.e. small grain and safflower proportions) can be estimated. These variables need not be input directly. A special FORTRAN subroutine is used

to transform the input variables into the variables of interest. This is convenient for a project where different measurement procedures may be used as input (i.e. dots counted, grams weighed) and changed to proportions within the program. Modifications to the original MPHASE allowed the generation of ratio estimators and the use of variable cluster sizes, weighting the proportions appropriately.

Specification of the Stratification Scheme

Based on the results of the Monte Carlo analysis (Section 4.1.3), the stratification scheme used for this year's statewide estimation was modified (See Section 2.2 for a description of the original stratification). The modifications were designed to reduce sampling variance as well as control measurement error. These new strata were composed of areas that, on Landsat 1:1,000,000 color composite transparencies, appear to be:

Table 4-6. Stratification scheme used in the allocation of sample units for the Task I estimation of irrigated land.

Stratum Number	Stratum Description
1	Generally dry farmed
2	Field crop areas dominated by fields less than 16 hectares (40 acres) in size
3	Field crop areas dominated by fields less than 16 hectares (40 acres) in size with known high proportion irrigated
4	Field crops dominated by fields 16 hectares (40 acres) or larger in size
5	Orchards and vineyards less than 16 hectares (40 acres) in size
6	Orchards and vineyards 16 hectares (40 acres) or larger in size
7	Unusual agricultural areas

The procedure used to produce the state-wide stratification is described in Section 4.2, Stratification and Sample Frame Construction. This revised stratification scheme will be evaluated at the end of the Task I inventory.

Specification of the Sampling Frame

For geographic areas, sampling frames usually are constructed as either a point system referenced by coordinates or an arbitrary clustering of areas into some convenient size unit (e.g. rectangular areas). The project objective as well as statistical and implementation considerations all enter into decisions which lead to the "optimum" strategy for sampling the population. Photo-related variables were (and may be in the future) a major part of the system either as a separate phase or as an aid to ground data collection. Therefore, the sampling frame should allow maximum use of the photographic capabilities for a given expenditure of effort. For this reason, point systems are not practical; to photograph a large number of different points with a single or pair of images is very costly. A cluster system is more economical since larger units allow additional information to be obtained at little incremental cost.

Initially, the decisions on sample unit size and configuration were based largely on practical considerations as insufficient data existed to simulate and optimize sample unit dimensions for large area inventories in California. A nominal 1.6 x 8.0 km (1 x 5 mi) sampling unit was used for both the 10-County (Section 2.1) and the 14-County (Section 2.2) studies because: (1) DWR's standard aerial survey photography covers a one-mile wide strip, (2) a five mile length is easily located and flown over several dates, and (3) the north-south orientation corresponds to DWR's survey techniques.

These same considerations were valid for the present study; thus, the nominal 1.6 x 8.0 km (1 x 5 mi) north-south oriented unit was maintained. Two modifications were made, however. Given the choice during sample frame development of having two small or one large sample unit, the larger unit was favored. This was done to decrease the errors due to possible misregistration of units when they transferred onto maps and Landsat enlargements. The second change was in sample unit orientation. The north-south orientation was maintained in the Central Valley and other agricultural areas where road networks were primarily oriented north-south. The sample units in upland areas and small valleys were oriented along major landforms and/or main thoroughfares. This was done to prevent having a large number of small sample units at the expense of having only very few large units, and increasing driving efficiency for the ground data collection.

Alternatives to the 1.6 x 8.0 kilometer sample unit size have been proposed by research cooperators at UC Santa Barbara and NASA-Ames Research Center. When addressing modifications to the area of a given sample unit, many interrelated variables must be addressed. In the absence of a large data set, one can make certain assumptions how variance and correlation change with sample unit size and how costs vary for ground data collection, aerial photography and Landsat acquisition, and data interpretation and measurement. Once reasonable assumptions are established, projections can be made as to the effect of changing sample unit size on estimate accuracy.

Using a set of reasonable assumptions, personnel at UC Santa Barbara compared the standard 1.6 x 8.0 km (1 x 5 mi) sample unit size to transect sample units approximately 1.6 x 36.3 km (1 x 22.5 mi). They state that in terms of

total flight time, "transect sampling may be a cost effective alternative to random segment sampling." (See Appendix I)

Using another set of reasonable assumptions, Arno (1979) compared five different sample unit sizes: 0.004, 0.040, 0.405, 2.590, 12.950 km^2 (1,10,100 640, 3200 acres). He states that "for a given cost, accuracy increases as the unit size increases up to 0.405 km^2 (100 acres). It peaks between 0.405 and 2.590 km^2 (100-640 acres), and accuracy decreases as the size grows to 12.950 km^2 (3200 acres)."

Further analysis by UC Berkeley personnel of these two separate investigations has indicated any size of unit can be justified based on a given set of reasonable assumptions. Depending on population size (N), correlation between Landsat and ground measurement (ρ), cost of Landsat and ground unit measurement (C_L and C_g, respectively), variance (σ^2, which varies with sample unit size), and desired accuracy (α), a convincing case can be made for designs ranging from an SRS with one acre sample units to a regression approach using long transects.

With the large data base from this 1979 Task I inventory, reasonable estimates and ranges for N, ρ, C_L and C_g, and σ^2 can be made under various accuracy constraints to determine the best sampling scheme for future surveys. The Task I Evaluation will address these very issues at the conclusion of the Task I inventory. Until then, the 1.6 x 8.0 km (1 x 5 mi) sample unit will be retained as it has proven to be workable.

4.1.5 Sample Allocation Computation

As can be seen in the analysis flow (Figure 4-2), the sample allocation computation was based on input from two major sources: (1) the specification of the mathematical model, stratification scheme and sample frame and (2) a sample unit list summarized by stratum and county.

Since the sampling design for Task I included the use of stratification, allocating the sample units required the distribution of sample units among the strata for each hydrologic basin. The distribution of units could have been simply proportional to the relative size of each stratum. Since the 1976 14-County Study gave estimates of within stratum variance (σ^2) and correlation (ρ), the optimum (theoretically giving smallest variance) allocation of sample units to each stratum (n_i) can be accomplished by minimizing variance subject to a cost constraint, as follows:

minimize: $$V(Y) = \sum_{i=1}^{\lambda} W_i^2 \, (1 - \rho_i^2) \, \sigma_{y_i}^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \left(1 + \frac{1}{n_i - 3}\right) \quad (16)$$

subject to: $$M = \sum_{i=1}^{\lambda} n_i \, c_i \quad (17)$$

where:

$V(Y)$ = estimate of variance of the estimate of basin proportion irrigated

λ = number of strata

i = stratum number

W_i = proportion of basin composed of stratum i

ρ_i = sample correlation between LANDSAT and ground data in stratum i as determined in the 14-County Study

$\sigma_{y_i}^2$ = sample variance of proportion irrigated for ground data in stratum i as determined in the 14-County Study

n_i = sample size in stratum i

N_i = population size in stratum i

M = maximum relative cost permitted in basin

c_i = weighted average relative cost of stratum i

The values of λ, W_i and N_i came from summary tables for each hydrologic basin (Table 4-7). The basin summary tables were compiled from similar tables constructed for each County (Table 4-8). The information summarized on the county table was derived from detailed county sample unit lists that described each sample unit in terms of agricultural practice stratum, presence or absence of grain and/or vegetables and relative ease of ground access (Table 4-9).

The constraint function (equation 17) uses the average relative cost of ground checking a sample unit in a particular stratum (c_i). In the 14-County Study all sample units (SUs) were located on the floor of the Sacramento Valley and were considered equally accessible. As sample units were allocated over the entire state for the 1979 inventory, the assumption of equal accessibility was not valid. Therefore, sample units were divided into three accessibility categories. Relative cost weights (c_i) were then determined for each stratum. Appendix II describes the development and use of the ground accessibility cat-

Basin: Tulare

County: All

Table 4-7. Example of a hydrologic basin summary table. The number of strata, proportion of the basin composed of each stratum and the population size per stratum came from tables like this summarized for each hydrologic basin.

VEGETABLE		1	2	3	4	5	6	7	
GRAIN	1				159				159
	2				85				85
	3				25				25
	D								
OTHER					150		20		170
					419		20		439

NON VEG		1	2	3	4	5	6	7	
GRAIN	1				201				201
	2				190				190
	3				46				46
	D	45							45
OTHER					438	58	171		667
		45			875	58	171		1149

			1	2	3	4	5	6	7	
ACCESS	A	1.00	42			1294	58	191		1585
	B	1.14	3							3
	C	1.38								
			45			1294	58	191		1588
			1.01			1.00	1.00	1.00		1.00 76.02

Basin: Tulare (TB)

County: Kern (KE)

Table 4-8. Kern County is shown as an example of the county summary table. Within each hydrologic basin (Tulare in this case) all counties were summarized in this way.

VEGETABLE		1	2	3	4	5	6	7	
GRAIN	1				32				32
	2								
	3								
	D								
OTHER					119		20		139
					151		20		171

NON VEG		1	2	3	4	5	6	7	
GRAIN	1				70				70
	2				55				55
	3								
	D	5							5
OTHER					176		13		189
		5			301		13		319

		1	2	3	4	5	6	7	
ACCESS	A	5			452		33		490
	B								
	C								
		5			452		33		490

Table 4-9. Example of a sample unit list generated for each county within each hydrologic basin. Each sample unit is described by agricultural practice stratum, grain and/or vegetables and accessibility. Sample units that were ultimately selected for ground checking were marked with an "E" (indicating a ground visit early in the season) or a ✓.

Basin: Tulare

County: Kern

Basin	County	Stratum	Grain	Veg	Access	Selection	#
		4	1	V	A	E	281
							282
							283
							284
		4	2		A		285
							286
							287
							288
							289
							290
		6		V	A		291
						E	292
							293
							294
							295
							296
		4		V	A		297
							298
		6		V	A		299
							300
							301
							302
							303
							304
							305
		4		V	A		306
							307
							308

egories to predict c_i.

After all terms were defined, a computer algorithm, FCDPAK*, was used to minimize Equation 16 subject to Equation 17.

For each hydrologic basin, the total number of sample units was allowed to vary over the range of 30 to 200. FCDPAK determined the optimal allocation of these units to each stratum (n_i). Percent confidence intervals at 95, 98, 99 and 99.9% levels of confidence were also calculated for each allocation. These percent standard confidence intervals were then plotted against the total sample size. Figure 4-5 illustrates a typical plot using the Tulare Basin allocation. From these plots, the total number of sample units required to achieve ±3½% at the 95% confidence level was determined by interpolation. As the stated inventory accuracy objective was ±5% @ 95% confidence level, this more conservative criteria insured against the possibility that chance alone would cause a failure to meet the stated goal in any hydrologic basin. As seen in Figure 4-5, the total number of sample units required to meet the ±3½% at the 95 criterion in the Tulare Basin was interpolated to be 65. This value is then compared to the FCDPAK values bordering this interpolated estimate (i.e. 62 and 81 sample units). The FCDPAK stratum allocation within the Tulare Basin for the 62 and 81 sample units is tabulated in Table 4-10.

To achieve the desired stratum-level allocation of the 65 basin units, a second interpolation was performed using the optimal FCDPAK stratum allocation for 62 and 81 basin units. This procedure was used for all the hydrologic basins. The resulting allocation of sample units by basin and by stratum is given in Table 4-11.

After all the sample units were allocated by stratum for each of the hydrologic basins, the units were physically annotated on map sheets for subsequent ground survey by DWR personnel. Measurement of both the sample units on the ground and the Landsat census is described in the following Sections.

4.1.6 Summary

The design process is a critical element in any inventory activity. It serves to specify the framework for data acquisition, analysis, summary, and storage and retrieval. By specifying this framework, all phases of an inventory are performed in a coordinated fashion, thus increasing the probability of successfully achieving the stated inventory objectives. For the design

---

*FCDPAK (Feasible Conjugate Direction Package for the Solution of Differentiable Mathematical Programs) was developed by Best (1974) to solve the general problem of maximizing a function subject to linear and/or nonlinear constraint functions. The program's only shortcoming is that solutions to n_i are generated in noninteger form. This problem was solved by use of the following contingency table:

$$\text{if } n_i - \text{integer}(n_i) < 0.1, \text{ then } n_i = \text{integer}(n_i)$$

$$\text{if } n_i - \text{integer}(n_i) \geq 0.1, \text{ then } n_i = \text{integer}(n_i) + 1$$

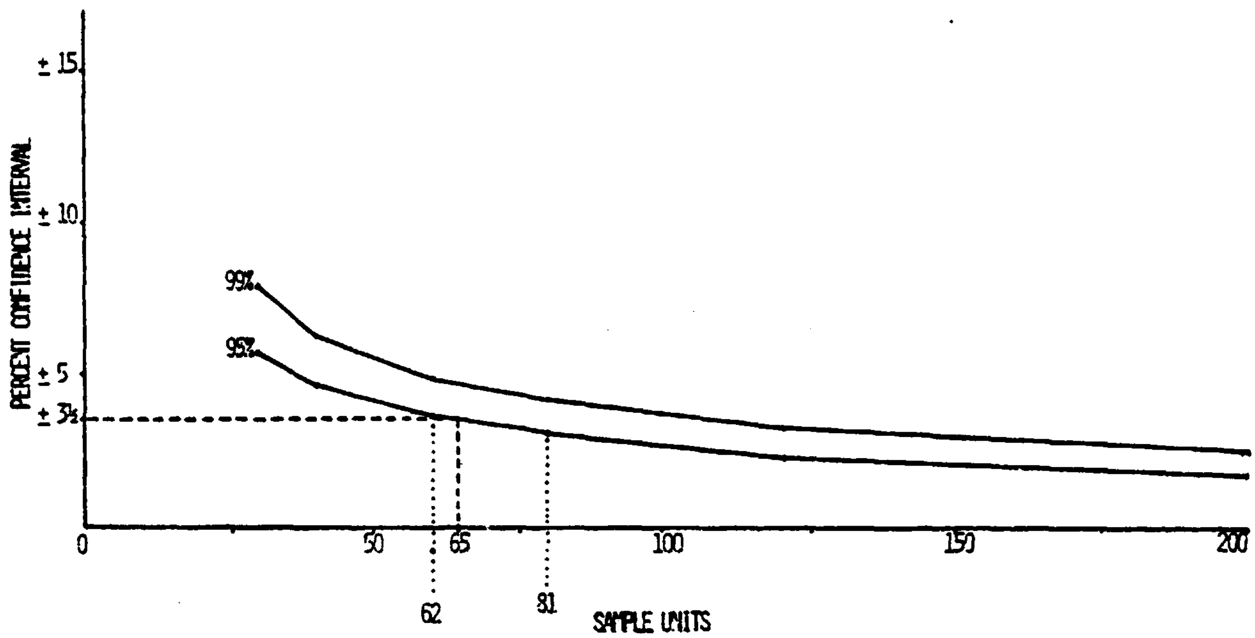


Figure 4-5. Percent standard confidence interval plotted against total sample size for the Tulare Basin.

Table 4-10. Allocation of sample units by stratum. The values were calculated by interpolation from the allocation shown in Figure 4-5.

Stratum	Values from FCDPAK	Values used by interpolation	Values from FCDPAK
1	4	4	4
2	0	0	0
3	0	0	0
4	44	46	61
5	5	5	5
6	9	10	11
7	0	0	0
Total	62	65	81
Accuracy	±3.6 @ 95	±3.5 @ 95	±3.1 @ 95

Table 4-11. Sample unit allocation by stratum for each hydrologic basin.

HYDROLOGIC BASIN	STRATUM 1	2	3	4	5	6	7	TOTAL
Central Coastal	26	7	28	8	-	5	6	80
Colorado Desert	-	-	-	42	4	8	4	58
North Coastal	4	6	-	30	12	-	-	52
North Lahontan	-	12	-	26	-	-	-	38
Sacramento Valley	8	10	-	39	5	4	6	72
San Francisco Bay	19	4	11	7	14	-	-	55
San Joaquin	6	5	-	53	5	14	-	83
South Coastal	12	18	9	11	24	-	8	82
South Lahontan	7	-	-	39	-	-	6	52
Tulare	4	-	-	46	5	10	-	65
California	86	62	48	301	69	41	30	637

process to be successful, the interrelationships between data collection, decision-making, and management must be understood, documented and integrated into the design. Only by understanding these interrelationships in concert with historical inventory practices can realistic assumptions be made, functional relationships documented, and operational systems developed and implemented.

The current design effort for the 1979 inventory had the benefit of close cooperation with the user agency, California Department of Water Resources, who provided invaluable management and decision-making insight and critical information needs on a state-wide basis. Furthermore, DWR conducted the complete ground survey effort providing the sensitive, costly, and compulsory ground data to drive the state-wide estimation process.

Based on both DWR input and the 10- and 14-County Studies conducted by the University of California, important historical data were available for the design process. The experience and the data were used to (1) generate and refine assumptions, (2) evaluate various estimator alternatives, (3) evaluate the effect of stratification on sampling and measurement errors, (4) calculate estimates of variance and data plane (i.e. Landsat-ground) correlations, both critical for the sample size calculations, and (5) generate accessibility/cost constraint functions paramount in the sample allocation process.

After numerous analyses, the inventory design was completed and implemented. The 1979 design may be summarized as follows:

GOAL: Estimate the proportion of irrigated acreage in the state of California to within ±5% allowable error at the 95% level of confidence.

DATA TYPES:

- Multitemporal Landsat color composite imagery enlarged to a scale of 1:150,000 (Phase I)
- Ground data collected by DWR; supplemented with 35mm aerial photography (Phase II)
- USGS maps at scales 1:1,000,000, 1:250,000, 1:62,500 and 1:24,000
- U-2 color infrared aerial photography at a scale of 1:130,000 and 1:24,000

SAMPLING FRAME:

- Sampling frame of area units (clusters)
- 1.6 x 8.0 km rectangular sample unit
- Orientation of sample units predominately north-south; allowed to vary with local topography and road network

STRATIFICATION:	• Hydrologic basin, county • Agricultural practice/land use • Small grain and vegetable • Exclusions
MATHEMATICAL MODEL AND SAMPLE ALLOCATION :	• Multiphase design • Census at Phase I (Landsat) • Simple Random Sample within strata/basin at Phase II (ground data) • Phases linked using regression estimator for large sample sizes and an unbiased ratio estimator for small sample sizes

When the statewide inventory is completed, a detailed evaluation of the Task I design process can begin. The evaluation will allow further refinement of assumptions, sampling frame (including size, shape and orientation of SU's), two phase sampling, stratification, sample allocation, and the estimation procedure (i.e. equation used to link phases and predict errors; and, procedures used to aggregate strata estimates into final estimates).

4.2 STRATIFICATION AND SAMPLE FRAME CONSTRUCTION

Stratification is a commonly used technique designed to reduce variance by systematically placing boundaries that separate homogeneous units. For Task I the major purposes of stratification were to:(1) allow summary of data by administrative units (hydrologic basin, county and state), (2) reduce sampling and measurement error, (3) enhance the allocation of sample units, and (4) flag areas for early and/or multiple ground data collection. The production of three stratifications was necessary to address the purposes just described: (1) administrative boundaries were defined by use of a DWR-supplied map delineating hydrologic basins and county boundaries were located from USGS 1:24,000 and 1:250,000 scale topographic maps (Figure 4-6); (2) an agricultural practice stratification was developed to reduce sampling and measurement error and enhance the allocation of sample units; and (3) areas of small grain and vegetable cultivation were stratified to help optimize ground data collection. The latter two stratifications will be described in Sections 4.2.2 and 4.2.3. As shown on the analysis flow (Figure 4-1), a merged stratification was formed that became the basis for the sample unit list required to compute the sample allocation.

Figure 4-6. Counties and hydrologic basins of California.

4.2.1 Regionalization of the State of California

The University of California, Santa Barbara has involved itself with items that impact stratification and the subsequent allocation of sample units. This has been in response to U. C. Berkeley's work in sample design.

Our work in stratification began with the definition of 12 photomorphic regions in the state based on three criteria: fog/cloud cover; target-to-background contrast; and presence/absence of agricultural activities. Regionalization defines those areas where remote sensing techniques are applicable to the task (e.g. 4 interior regions where satellite remote sensing can be used were noted as well as 5 regions without agriculture and 3 coastal regions where fog will most likely interfere with data acquisition) and those areas where similar techniques can be used.

Subsequent work lead to the definition of subregions, or clusters of counties with similar crop mixes. While this information may have its greatest value in Tasks III and IV, it was useful for defining those counties with "problem crops" such as grains and vegetables that must be considered in Tasks I and II.

This effort was supported by two questionnaires sent to the U. C. Cooperative Extension Office in each county. The first questionnaire was concerned with the acreage, the timing, and the specific crops involved in multicropping (i.e. double or triple cropping). Approximately 75% of the counties have responded. The second questionnaire was concerned with small grains - the amount of irrigated *vs*. non-irrigated grains, the specific grains involved and cropping practices. Approximately 30% of the counties responded. The general pattern seen in the responding counties is that most grains are not irrigated regularly. Irrigation most often occurs during the preparation stage and occasionally once between emergence and maturity. This is highly variable from year to year, depending on seasonal rainfall conditions.

Utilizing the work in subregionalization as well as the questionnaire results for multicropping and Landsat color transparencies, multicropping strata were defined for the purpose of allocating samples that require early or late field visits to detect second crops. This work was done in close cooperation with U. C. Berkeley and will be used in addition to their earlier stratification scheme based on field size and crop type (field crop *vs*. orchard).

Much of the work done on stratification for Task I will be extended to Tasks III and IV. One possibility is the use of DWR land use summaries for 7.5' quandrangles to define crop mix strata for sample allocation and possible *a priori* classification in Task IV. The data is already in a computer compatible format and could be tested for a small region. It possibly could be useful for defining strata based on the amount of irrigated acreage for Task II. While this requires certain assumptions about crop stability over time, the DWR land use surveys represent some of the best data available.

Two other sources of data that may prove valuable for stratification, especially Tasks III and IV, are crop maps produced by the Soil Conservation Service Statewide Important Farmland Mapping Program and county crop maps

found in the California Department of Food and Agriculture's Report on Environmental Assessment of Pesticide Regulatory Programs. These data sources should be examined for strata definition for crop identification.

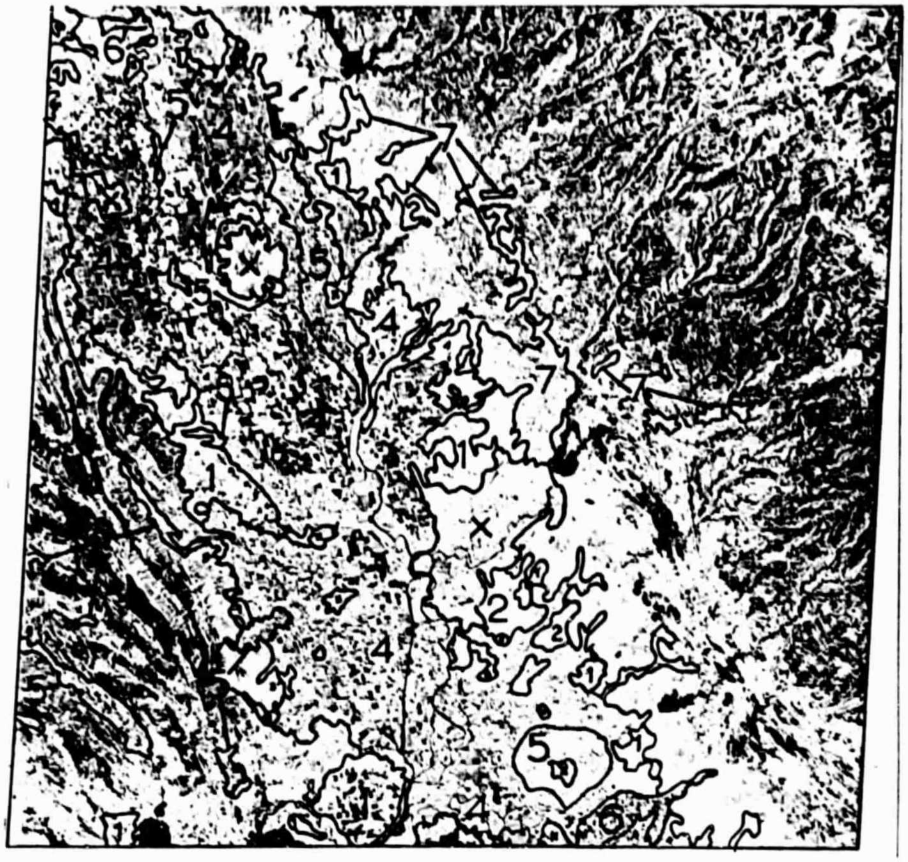
4.2.2 Agricultural Practice Stratification

The agricultural practice stratification developed for the Sacramento Valley (Section 2.2) was used as the base for the 1979 work. It was based on two general factors that are critical to both manual and digital classification of Landsat; land use and field size. When defining land use for the purpose of estimating irrigated agriculture, there are several pertinent factors to be examined: (1) the presence/absence of any agriculture; (2) historically known or topographically defined areas of dryland vs. irrigated agriculture and (3) variations in agricultural cropping practices within a generally irrigated area (i.e. field crops vs. orchards). The problems caused by small field size affect the human analyst where detecting and identifying fields as well as accurately drawing boundaries becomes difficult and tedious and to the computer where the edge effect of mixed pixels and precise registration of acquisitions is critical. Before extending the stratifications completed on the Sacramento Valley to the rest of the state, Monte Carlo tests were performed examining the value of the strata (Section 4.1.3). Based on the results of the tests, a modification of the original stratification scheme was used (Table 4-6).

To minimize interpreter variability, the entire state (approximately 30 Landsat frames per date) was stratified by a single analyst into one of the seven strata described above. Since the minimum sample unit size was one square mile, areas less than that were not delineated (areas less than one square mile are subject to measurement for total irrigated acreage on Landsat but were considered too small to act as individual sample units). Stratification was done by overlaying clear acetate on 1:1,000,000 Landsat color composite transparencies and delineating the appropriate stratum. Multitemporal Landsat imagery was used to verify the consistency of the delineation. Since quite different agricultural practices and, therefore, quite different strata may appear similarly on any single date of imagery, it is very important to utilize the multitemporal capability and synoptic coverage of full frame Landsat to obtain an accurate, repeatable stratification. California's virtually cloud free summer growing season over the major agricultural areas lends itself particularly well to the availability of a large set of Landsat data for this purpose. Figure 4-7 shows an example of the agricultural practice stratification used in the Sacramento Valley.

4.2.3 Small Grain and Vegetable Stratification

In order to direct the collection of field data two additional stratifications were necessary. Areas of small grain and vegetable cultivation have historically posed a problem in ground data collection due to: (1) early harvest of grains and subsequent plowdown, and (2) multiple cropping in vegetable areas (Section 6.1 describes the dynamics of multiple cropping in the south coastal



Stratum Number	Stratum Description
1	Generally dry farmed
2	Field crop areas dominated by fields less than 16 hectares (40 acres) in size
3	Field crop areas dominated by fields less than 16 hectares (40 acres) in size with known high proportion irrigated
4	Field crops dominated by fields 16 hectares (40 acres) or larger in size
5	Orchards and vineyards less than 16 hectares (40 acres) in size
6	Orchards and vineyards 16 hectares (40 acres) or larger in size
7	Unusual agricultural areas

Figure 4-7. Agricultural practice stratification. Stratification similar to this was completed for the entire state and was used for the allocation of sample units that were ground checked. (Sacramento is located slightly southeast of center and marked with an "X").

area of California). To ensure ground data acquisition at the optimum time, areas of grain cultivation and vegetable cultivation were stratified separately on 1:1,000,000 Landsat transparencies for each hydrologic basin.

After examining historical data on vegetable cultivation, boundaries of historical vegetable cultivation areas published by the California Crop and Livestock Reporting Service were transferred to the Landsat imagery. These boundaries were refined by reference to the Landsat imagery to account for land use changes and urban encroachment.

Small grain cultivation areas were delineated through analysis of 1976 through 1978 Landsat imagery. The grain areas were then classified into: (1) dryland grain farming; (2) areas of less than 21 percent grain; (3) 21-40 percent grain; and (4) greater than 40 percent grain.

Areas where multiple cropping occurs were examined through historical data and information from the county farm advisors. (Regionalization, Section 4.2.1) Most multiple cropping in California occurs in grain areas, where grain is followed by a field crop such as corn or beans, or in vegetable areas, where one vegetable crop follows another. The previous delineation of the grain and vegetable cultivation areas, therefore, included the majority of the multiple crop areas.

4.2.4 Formation of the Merged Stratification

Merging the agricultural practice, small grain and vegetable stratifications as well as locating administrative and exclusion areas was necessary before the sample unit list could be generated. Locating administrative boundaries, such as counties, exlusion areas (established wildlife refuges, cities), and assigning an access code (Appendix II) is facilitated by reference to available maps. Since the agricultural practice and croptype stratifications were based on the spatial and spectral information provided by Landsat, it was felt that an appropriate base for the merging of these functions was a combination of 1:250,000 scale USGS topographic maps and 1:250,000 scale Landsat enlargements. Enlargements were made on a county basis, by reference to the USGS maps. These enlargements and the associated maps provided the base upon which the sample frame of 1.6x.8 km (1x5 mile) units was created. The subsequent sample unit lists provided the population from which the ground data units were selected. In addition to providing the sample frame base, the combination of information available from the maps and enlargements was critical for accurate transfer of the sample unit boundaries selected for ground checking to the 1:24,000-scale (7.5') USGS maps used by DWR for field work.

For each of the 58 counties in California, the land use strata, grain cultivation boundaries, and vegetable cultivation boundaries were enlarged from the orginal 1:1,000,000 scale to the 1:250,000 scale Landsat prints. Using an overhead projector system, each boundary was projected, scale matched to the enlargement and drawn on a clear film overlay. By matching topographic features on both the original transparencies and the enlarged prints, accurate transfers

of boundaries were made. In situations where two or more strata had boundaries that were approximately coincident, they were merged into a single boundary (Figure 4-8). Color coding allowed for the differentiation of land use, grain and vegetable strata. Once the enlargement was complete, the overlays could be used on either the Landsat prints or the 1:250,000 scale topographic maps.

County boundaries were drawn from the 1:250,000 scale maps and overlayed on the merged strata boundaries. At this point all image defined agricultural phenomena were tied to the county base map. Hydrologic basin boundaries, provided by DWR, were transferred onto the overlays for those counties that were split into more than one hydrologic basin. Accurate location of this boundary was particularly important in those areas where the basin boundary crossed agricultural land, since misplacement of the boundary would result in farmland being transferred to the wrong basin.

4.2.5 Generation of Sample Unit List

When the merging of the strata and the location of county, hydrologic and exclusion areas was complete, each county consisted of a set of irregularly shaped polygons defined by some combination of the strata. Each polygon was labelled indicating the appropriate land use stratum, the presence of vegetables, the presence and proportion of grain and the general accessibility of each polygon. The merged and annotated overlay was then placed over a gridded template of 1.6x8 km (1x5 mi) sample units and the 1:250,000 USGS map. In those areas where the predominant field pattern was oriented north-south, the sample unit grid was placed to coincide with section lines. This was done to increase the ease and efficiency of the field data collection effort. In areas where the topography or historical land development caused the dominant field pattern to be oriented in other directions (i.e. Salinas Valley) the sample unit grid was placed so as to conform with the developed road/field pattern system. The sample unit grid was traced onto the county boundary overlay for all areas that fell within the stratified area.

Although a sample unit was nominally defined to be 8 km (5 mi) long, actual length varied from 1.6 to 11.2 kilometers (1 to 7 mi). Editing of the sample units removed those less than 259 hectares (640 acres) in area and those portions of units that were less than .4 kilometer (.25 mile) wide. Each sample unit was then numbered and placed in a sample unit list.

The information from the sample unit list was summarized in a table for each county. Similar summary tables were made for each hydrologic basin. The basin summary sheet was used to calculate average relative access cost within each stratum and the proportion of the basin represented by each stratum. This information, along with the number of sample units in each stratum (the population size) was used to compute the stratum sample sizes as described in Section 4.1.5.

4.2.6 Preparation of 7.5 Minute Quads for Ground Measurement

After the units to be ground surveyed were randomly selected, the boundaries of these sample units were visually transferred from the 1:250,000 scale overlay to 1:24,000 scale overlays. Standard DWR procedures call for the use of USGS 7.5' set of maps and contained the selected sample units. Field crews from DWR used these overlays for field mapping of irrigated crop land.

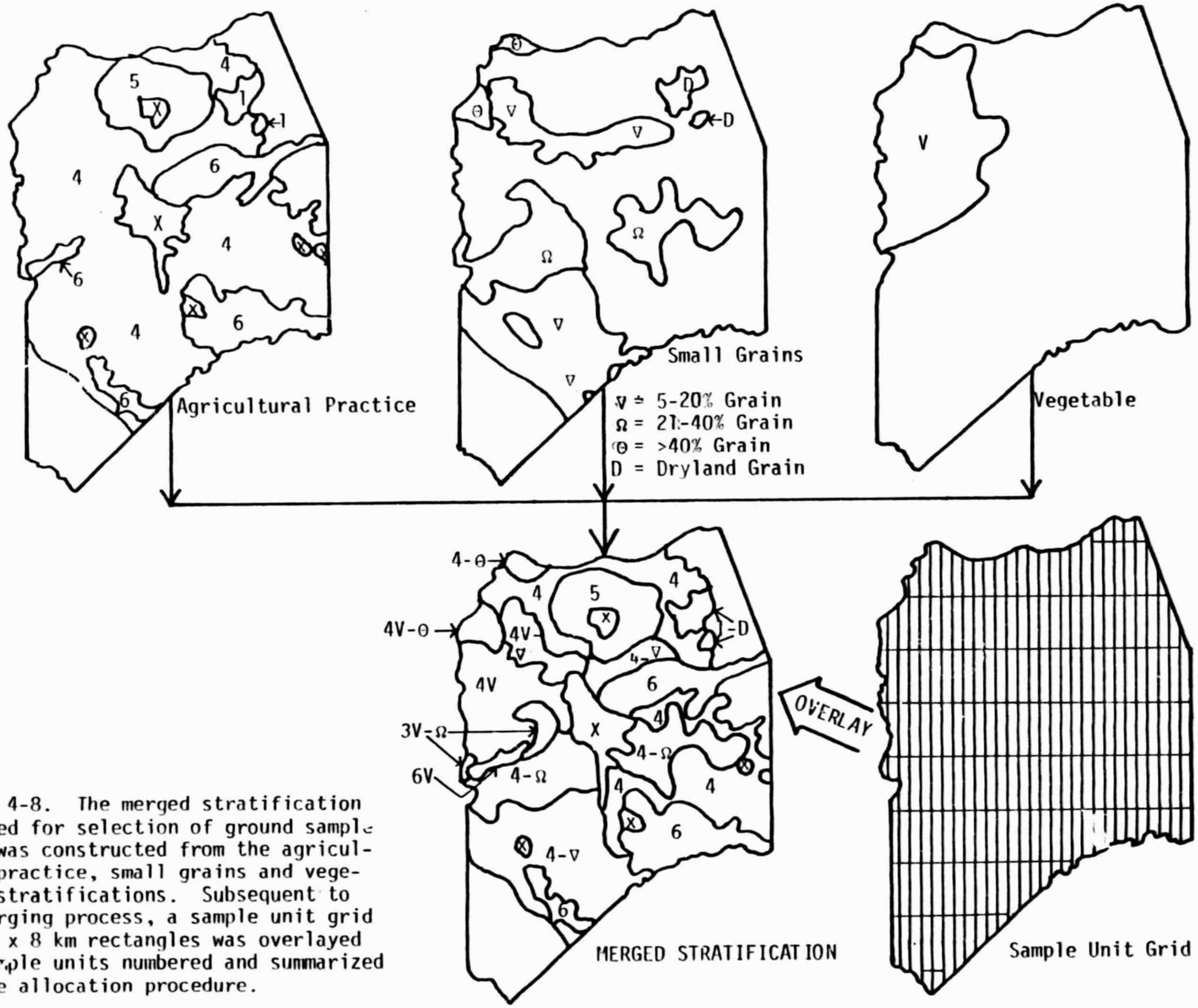


Figure 4-8. The merged stratification required for selection of ground sample units was constructed from the agricultural practice, small grains and vegetable stratifications. Subsequent to the merging process, a sample unit grid of 1.6 x 8 km rectangles was overlayed and sample units numbered and summarized for the allocation procedure.

Figure 4-10. Center points of the thirty-three Landsat scenes needed to image California.

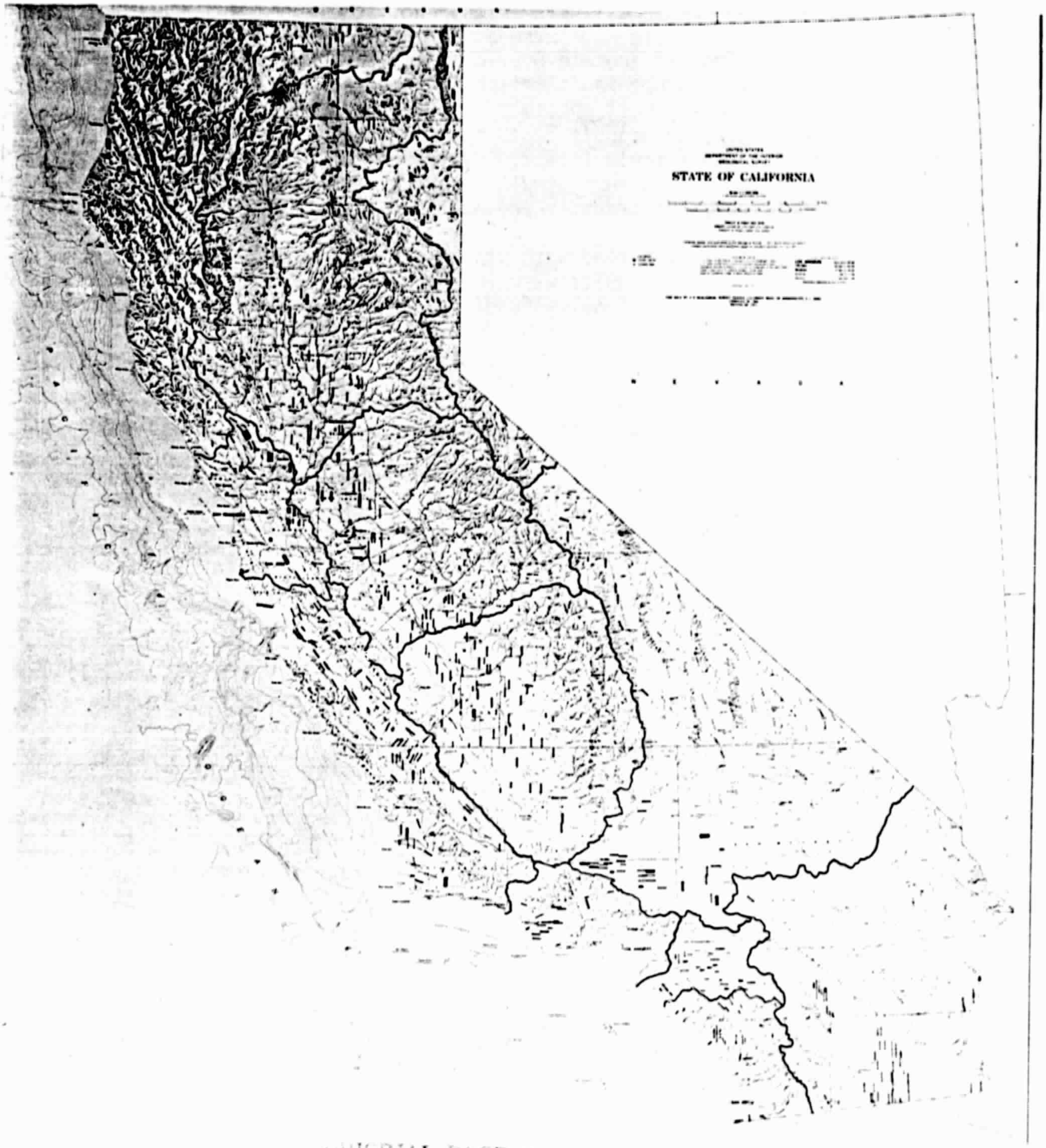


Figure 4-9. Distribution of ground (Phase II) sample units. Each of these 637 units was checked by DWR to determine the location of irrigated fields.

Figure 4-10. Center points of the thirty-three Landsat scenes needed to image California.

Normally, California's virtually cloud-free summers over major agricultural areas provide ample opportunity to select from a large variety of acquisitions to obtain the optimal data set. In 1979 the satellite and ground processing problems often combined to severely limit or nullify any choice of acquisitions. Appendix III lists the acquisitions used for each county. Although certainly not the optimal date selection, three time periods of imagery were generally available for each of the counties.

4.3.2 Enlarging and Mosaicing Landsat Frames

In 1979, as in the earlier projects, measurement at the Landsat phase is done on 1:150,000 scale enlargements of each county. On a county basis, each available Landsat frame was evaluated for image quality (i.e. line drop, "smearing"), color balance, exposure and miscellaneous items such as cloud and smoke. Following this evaluation, the best combination of dates and frames was selected for enlargement.

In order to maximize the efficiency of the darkroom work, each transparency selected for enlargement was prepared as follows:

- 1:1,000,000 scale county boundaries from a USGS map were overlayed on the transparencies
- Templates were prepared that outlined the area that would be covered by 8x10, 11x14, 16x20 and 20x24 inch photographic prints
- The appropriate combination of templates needed to photograph the county was selected, annotated and numbered on each transparency
- Ten mile segments were randomly measured over the area of the transparency. A scale (100 lines/centimeter) and an enlargement factor were then aligned on the border of each template and included when the negative was copied.

These prepared transparencies were sent to the darkroom for enlargement. Scale matching was done by use of: (1) the 100 lines/cm scale and enlargement factor on the transparency, (2) 1:150,000 scale county boundaries plotted by the Office of Geometronics of Caltrans (State of California, Department of Transportation) and (3) reference to 1:250,000 scale USGS maps for topographic features and location of the county boundaries.

At this time, sixteen of the fifty-eight counties in California have been enlarged. These counties represent approximately 33% of the total land area of California but contain 60% of the total possible agricultural sample units.

When the enlargement was completed, each county was mosaiced together and mounted on stiff posterboard. Counties that would have a mosaiced size greater than approximately 750 cm x1 meter were divided and mounted on separate boards. This size limitation facilitated handling and interpretation as well as storage.

4.3.3 Generation of Recording Forms

Forms for recording the interpretation done on the multitemporal Landsat enlargements were created for each county. To produce the form, the 1:150,000 scale county boundaries plotted by Caltrans were located on one of the completed mosaics for each county. The county boundary was then traced onto a second overlay; the originally plotted boundary was archived. The agricultural practice strata, exclusions and hydrologic basin boundaries were transferred from the 1:250.000 scale overlays by interpretation. The agricultural practice strata boundaries were necessary because (1) interpretation responsibilities were divided between analysts based on these strata boundaries, and (2) digitization of interpretation results was needed by stratum. Exclusion areas were also transferred from the 1:250,000 overlays; reference was also made to 1979, U-2, 1:130,000 scale CIR aerial photography to refine boundary placement. The hydrologic basin boundaries were needed for summarization of results and as a logical way to divide work between the Berkeley and Santa Barbara campuses. Superimposed on the overlay, which was now a composite of county, agricultural practice strata, exclusion and hydrologic basin boundaries was placed a grid that defined the borders of 7.5 minute quadrangles. The grid was used as a mechanism for organizing interpretation, a unit for documenting the time required to perform interpretation and as a potential area for summarization and comparison of results with DWR's land use surveys.

4.3.4 Interpretation of Multitemporal Landsat Imagery

The interpretation logic and procedures for identifying irrigated land in California are basically the same that have been used in the past projects. The analyst is required to make a decision on whether a particular parcel of land is irrigated. To do this the analyst relies on a variety of image characteristics and logical expectations of the presence and appearance of irrigated land.

Providing the analyst with sufficient data to develop reasonable expectations is critical to accurate measurement at the Landsat phase. Prior to interpreting a particular county the analyst is given a variety of ancillary information upon which to build his expectations. These include (1) California Crop-Weather which is published on a weekly basis by the California Crop and Livestock Reporting Service and summarizes weather conditions over the state as well as land preparation, planting, growth condition and harvesting of field crops, fruit and nut crops, vegetable crops and livestock (pasture and range conditions). The information is summarized by region and provides the means for constructing year/ gional specific crop calendars; (2) Agricultural Commissioner's crop reports for 1979 which list county acreage by crop type; (3) California-Arizona Farm Press that publishes weekly reports on all facets of agriculture in the West including land preparation planting, irrigation and water problems, pest and disease management, fertilization, plant variety performance, economic marketing and tax issues, legislation and harvesting; (4) California Grower and Rancher - a monthly published magazine on agriculture in California (written and published regionally); (5) 1979, U-2, 1:130,000, color infrared photography of the majority of agricultural land in California, and (6) antecedent DWR land use survey quads and summary statistics. Using all or a subset of the available data, the analyst builds a mental model of what he expects to see on the dates of Landsat imagery provided for each county.

Image characteristics traditionally used in manual photographic interpretation are exploited in the analysis of Landsat imagery. For the majority of the interpretation, the most critical characteristics are (1) pattern (is this an area of agricultural fields?) and (2) color (is this field the color expected for an irrigated field on the date being analyzed?). Other critical characteristics that analyst relies on are texture, shape of fields, and location of fields. These last three characteristics are particularly important when interpreting in mountain areas, along rivers and streams, (intermingled riparian vegetation) on the fringes of well developed agriculture and in areas of dispersed agriculture such as the foothills.

The interpretation procedure calls for analysis to be done in a specific manner. The structure of the interpretation system is designed to (1) eliminate variability in the method interpreters use and (2) allow for a detailed evaluation of the separate parts of the analysis system. The procedure calls for:

- Within each hydrologic basin assignment of a single interpreter is made to each stratum. An interpreter may analyze more than one stratum per basin, but no stratum should be interpreted by more than one analyst.

- Using the 7.5' grid as a base, interpretation proceeds on a quad-by-quad basis moving left to right and top to bottom

- Interpretation is done on the mid-summer date first, the spring image second and fall date last. In strata where irrigated agriculture dominates, the analyst delineates areas that are <u>not</u> showing active vegetative growth in July/August. These areas are marked with a single dot. The overlay is then placed over the May image and the blocks marked with a single dot are checked; if these areas are interpreted as irrigated cropland in May, a second dot is added. The analyst then proceeds to the final date and checks the remaining singlely-dotted areas.

- Within each 7.5' quad the analyst records the time required to interpret each stratum on each date

- Re-check areas as necessary

4.3.5 Digitization of Measurement Results

Upon completion of the interpretation, the results must be tabulated for input to MPHASE. The first step in this process is to locate the sample units that had been selected for ground checking. Accurate location is absolutely vital to the estimation procedure since the comparison of ground proportion irrigated to Landsat interpreted proportion irrigated "corrects" the estimate and provides the data needed to compute accuracy statements. Location is accomplished by reference to the 7.5' quadrangle maps with an overlay of the ground

annotated sample units. By visual comparison of the ground data (field pattern) and map features (i.e. roads, canals, railroads) to the 1:150,000 scale Landsat enlargements, accurate location is possible.

The proportion of irrigated land is then calculated by digitizing the total area of each sample unit and the area that is irrigated. Each sample unit is digitized and recorded separately. The remainder of the interpretation is digitized by stratum within each county.

4.3.6 Irrigated Cropland Mapping Procedure

Some water management applications require the use of spatially defined data as available only in map formats. As a small part of our effort, we have continued to improve and evaluate cropland mapping procedures in cooperation with Kern County Water Agency (KCWA). Since 1972, yearly maps of irrigated cropland has been supplied to KCWA based on manual interpretation of satellite or aircraft acquired imagery. More recently a multistage mapping procedure has been implemented to effectively integrate both types of imagery. Figure is a map generated from 1978 Landsat imagery and previous data provided by KCWA. An update of 1979 using the procedure shown in Figure 4-11 is underway.

A combined satellite and aircraft approach takes advantage of both the temporal frequency of Landsat multispectral imagery and the higher spatial resolution of aircraft photography to provide a product more useful than is available from either source individually. Multidate Landsat imagery is necessary for accurate mapping of irrigated cropland because of Kern County's long growing season, its numerous crops, and cropping practices (e.g., double cropping). While aircraft photography is obtained much less frequently than Landsat imagery, its higher spatial resolution is more suitable for detailed feature mapping (e.g., field boundaries and homesteads) and identifying specific ground conditions; this detailed information has proved to be highly complementary to temporal Landsat imagery in the classification process.

A comparison with California Department of Water Resources field based maps for approximately 175,000 acres (five 7½" USGS quadrangles) has shown the multistage approach to be very accurate, with only young permanent crops (e.g., orchards and vineyards) causing interpretation problems. Since permanent crops are relatively stable, once established, this problem appears amenable to improvements in the interpretation procedures; periodic ground surveys or high resolution aircraft photography should allow identification and mapping of permanent crops without yearly reinterpretation except to check for removal.

Since 1976 Kern County Water Agency has funded the yearly cropland updates. The one man-month effort required to complete each update provides a cost-effective and operational demonstration of remote sensing technology for water management purposes. Based upon the success of this ongoing program we plan to investigate the potential of a cooperative mapping program involving all of the water agencies in the Central Valley.

Future research activities include an evaluation of Landsat RBV imagery as a partial replacement for the aircraft photography and a digital implementation of the multistage procedure. It appears possible to effectively combine both mapping and sampling procedures into a system that can provide spatially defined products with quantitative confidence statements regarding total irrigated acreages.

CROPLAND MAPPING PROCEDURE

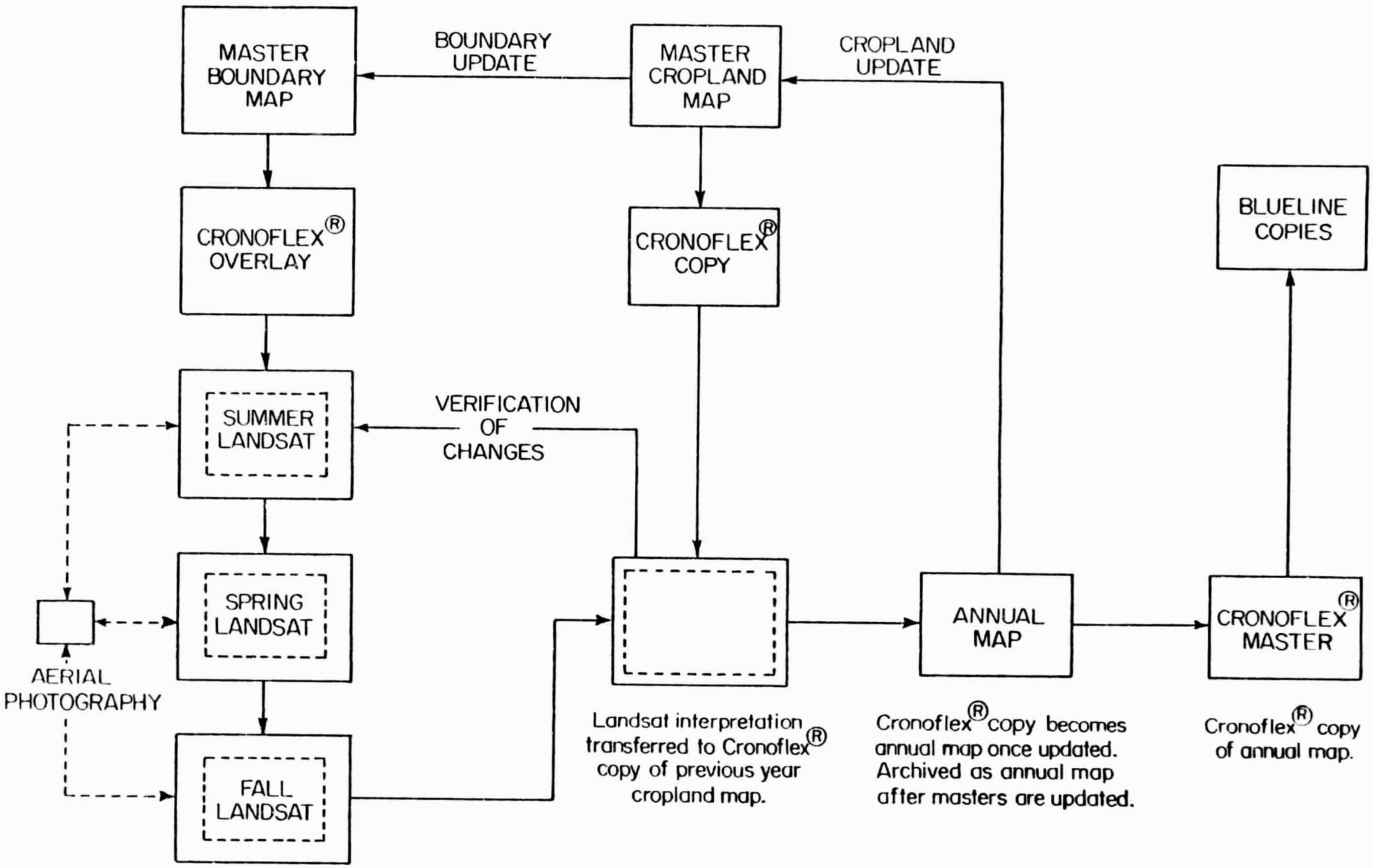


Figure 4-11

4.4 GROUND MEASUREMENT

For each of the 637 Phase II sample units, DWR district personnel made a field-by-field inspection to determine the presence of irrigation. Using 7.5' USGS quads with the plotted sample unit outlines as a base, (Section 4.2.6) field boundaries were drawn and each field coded. In many cases, detailed ground data including specific crop type mapping was done. At a minimum, the ground crews mapped parcels as irrigated or non-irrigated grain, safflower, field crop, pasture, other agricultural classes or lawn areas; fallow, farmsteads, feedlots, or dairies; native vegetation, water surfaces or unsegregated native vegetation; and six classes of urban. More than one visit was made to many of the units to verify multiple cropping.

When collecting the field data DWR often used their previously mapped land use survey quads (Figures 1-1 and 1-2) and the 35mm color aerial slides from which the maps were derived as aids for defining field boundaries. Color infrared 1:130,000 scale aerial photography flown by the U-2 during the spring and early summer of 1979 was also used extensively as soon as it was available. For a few units where access was particularly difficult, low altitude aerial observation of the unit provided the necessary information.

Each of these sample units was then tabulated by DWR and acreages output in a variety of forms: (1) by hydrologic basin - individual sample units listed by county (Figure 4-12); (2) by 7.5' quadrangles - sample unit(s) and county; (3) by county-cumulative summary of all sample units within the county; and (4) by DWR district-cumulative summary of all sample units mapped by the individual district offices. In total, DWR personnel ground checked (at least once) and tabulated approximately 520,400 hectares (1,286,000 A) across the state.

4.5 ESTIMATE SUMMARY, EVALUATION AND REPORT

The final estimate will be completed in mid-1980. At this time enlargement of Landsat, interpretation and tabulation are in full production. Following the calculation of the results, a detailed evaluation of the individual inventory system components and the overall system performance will be done. Working closely with DWR, the following system components will be evaluated:

- Sampling frame
    - Stratification
    - Sample unit size
    - Sample unit orientation
    - Sample unit frame construction
- Sample Allocation
    - Revised allocation using statistics available from 1979 data set
- Comparison of revised allocation and expected variance to allocation used and variance achieved

79-14	STATE OF CALIFORNIA THE RESOURCES AGENCY		CO. SERVICE AREA SUMMARY		LAND USE — IN ACRES — 1979		LAND SAT-SOUTHERN DIST		DEPARTMENT OF WATER RESOURCES		02/04/80

Unit Number	Co.	Service Area Code Number	Land Use Symbol Main	Special Condition	Irrigated Lands Including Fallow	Irrigated Lands Excluding Fallow	Nonirrigated Lands Including Fallow	Nonirrigated Lands Excluding Fallow	Service Area and/or Hydro Unit Totals	Quad Totals
	13	0002	G		227	227				
					227*	227*				
			R		235	235				
					235*	235*				
			F		44	44				
					44*	44*				
			P		378	378				
			P	T	81	81				
					459*	459*				
					965**	965**				
			NW				1852	1852		
							1852*	1852*		
			NC				772	772		
							772*	772*		
							2624**	2624**		
			U				123	123		
							123*	123*		
							123**	123**	3712	
	13	0004	G		167	167				
			G	T	111	111				
					278*	278*				
			F		352	352				
			F	T	317	317				
					669*	669*				
			T	T	33	33				
					33*	33*				
			I		156	156				
					156*	156*				
					1136**	1136**				
			NC				390	390		
							390*	390*		
							390**	390**		
			U				8	8		
							8*	8*		
							8**	8**	1534	
	13	0016	G	T	511	511				
					511*	511*				
			R		202	202				
					202*	202*				
			F		1166	1166				
			F	T	509	509				
					1675*	1675*				
			PI		87	87				
					87*	87*				

Figure 4-12. DWR tabulation of individual sample units by county.

- Comparison of sample variance from proportional versus optimal allocation of SUs for a set of fixed sample sizes

- Determination of sample allocation and resulting sampling variance from probability proportional to estimated size (PPES) SU selection

- Determination of possible cost savings using systematic selection of area SUs for given sample variance goals

- Measurement procedure

  - Landsat image interpretation

  - Ground data collection

  - Digitizing method

- Irrigated land area estimation procedure

  - Equations used to link sample phases to produce area estimates

  - Equations used to predict errors associated with area estimates

  - Procedures used to aggregate stratum estimates into final estimates

An evaluation of the overall system performance will also be done. This will include:

- Determination of estimate error by reporting unit relative to DWR baseline

- Determination of the sensitivity of the final error estimates to individual inventory components

- On the basis of the inventory results, development of expected error versus cost curves for given levels of statistical confidence

- Summarization of expected through-put rates

Following the detailed procedure outlined above, the results of the evaluation will be reviewed in concert with DWR and NASA to determine whether the inventory system demonstrated during 1979 met DWR's performance requirements at that time. Recommendation will be given as to what changes, if any, should be made before a future operational implementation of the inventory system.

5.0 ESTIMATION/MAPPING OF IRRIGATED LAND USING DIGITAL ANALYSIS TECHNIQUES (TASK II)

The digital analysis of Landsat multitemporal data for inventorying irrigated land received increasing emphasis in 1979. The information requirements for Task II were essentially the same as for the manual analysis of irrigated land. That is, the type of information needed is the estimation and/or mapping of irrigated land; the area of summary will eventually be the same as for Task I; and the performance criteria for the estimation procedure used ± 5% at the 95% level of confidence as a baseline.

Although representing only 20% of the total effort this year, significant progress was made on several sub-tasks. The sub-tasks were designed to address the major goals of 1979: (1) analyze and evaluate methodologies for the registration of multitemporal Landsat digital data and (2) test and evaluate various classification procedures. Three test sites were used: A 1° block in the Sacramento Valley studied by UCB for registration and classification procedures; and at UCSB, a 1° block and three 7.5' quadrangles testing registration and the same three 7.5' quadrangles evaluating classification methodologies (Figure 3-1).

5.1 REGISTRATION OF MULTITEMPORAL LANDSAT DIGITAL DATA

Because the precise registration of multitemporal Landsat digital data is imperative for accurate classification, significant effort was put on the exloration and testing of two major registration procedures:

- Control point least-squares analysis, and
- Cross correlation

5.1.1 Control Point Least-Squares Analysis - Remote Sensing Research Program (RSRP), UC Berkeley

In evaluating the candidate registration system, a number of questions were addressed:

- Could this system be efficiently used on a mini computer?
- What, if any, problems would be encountered when "sewing" adjacent Landsat paths together?
- How many control points are required to satisfactorily register and rotate to north multitemporal Landsat scenes?
- How could files be created that are based on USGS 7.5' quadrangles (DWR's standard map base)?

The test site selected for analyses was in the Sacramento Valley and consisted of a 1° block divided into four 30' segments. Each 30' block was a set of sixteen 7.5' quadrangles (Figures 5-1 and 3-1). The 30' block size was selected because: (1) it was convenient for storing and manipulating Landsat in a variety of forms for real time interaction, (2) coordinate transformations performed on an area this size were expected to maintain acceptable multitemporal registration accurate at the 15' and 7.5' quad size and (3) the 30' block is a multiple of the 7.5' quad which is DWR's standard reporting unit and which is

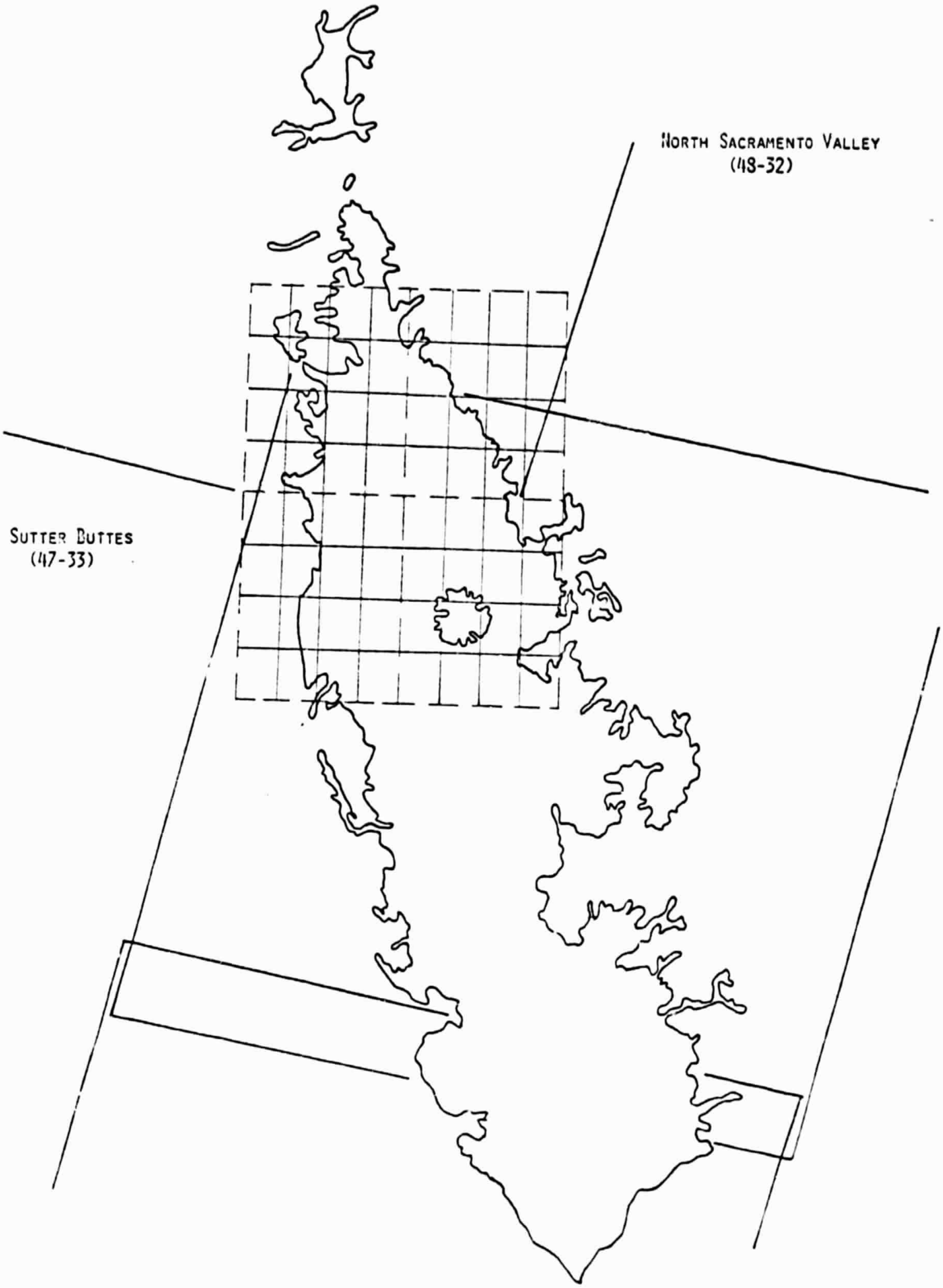


Figure 5-1. Location of the 1° block test site in the Sacramento Valley. The block was divided into four 30' segments. Each 30' segment was composed of 16 7.5' quadrangles.

always located completely on one Landsat scene.

The multitemporal Landsat data used as a test set for registration was also to be used for classification of irrigated land in the second sub-task. Date selection was controlled by a number of factors: (1) a three-date system, similar to that used in Task I, was to be tested; (2) the tes selected should, therefore, mimic the time windows used in a Task I system, and (3) 100% ground data for the test site was available for the 1976 growing season. The dates selected for registration and classification were:

Scene 47-33	Scene 48-32
30 May 76	22 May 76
28 August 76	20 August 76
3 October 76	4 October 76

The first step in the multitemporal registration of the Landsat data was to create, for each date, a data file on the disk containing an entire 30 minute block. Initially, the raw data was displayed from tape and point and line coordinates for the area containing the block were determined. This area, somewhat larger than the block itself in order to accomodate a north-south rotation, was then placed on the RSRP data disk. Because the block area usually required more than one quad of the Landsat scene, it was necessary to (1) create a file from each Landsat quad and (2) merge those files to create a new disk file containing the 30 minute area. To merge the Landsat files, a dummy file was created of the appropriate size with all values set at zero. The Landsat files were then transferred to the proper coordinates in the dummy file and the area was "sewn" together. This creating and merging was done for each of the three dates, late May, late August, and early October. An MSS 7/5 ratio band was then created for each date by multiplying the value in MSS 7 by 2 and dividing that product by the value in MSS 5 for each pixel.

A set of control points was selected to initiate registration of the multitemporal data set. One set of control points was used for the three dates. These points were distributed as evenly as possible over the 30' block with approximately three points per 7.5' quad area. Control point coordinates were obtained by displaying the disk file for each date, moving the cursor on the TV monitor to the selected point, and recording the x and y coordinates. Control points were selected based on (1) the case with which they could be located on the three dates of Landsat and the DWR ground data maps, and on (2) an approximately even distribution of points over the 30' block. The x and y coordinates of each control point were then measured on 7.5' quads in 1/60 inch increments. Measurements were made using the upper left corner of each 7.5' quadrangle as the origin.

Dimensions for the 30' north-south ground coordinate computer file were set at 620 points by 800 lines. These dimensions were chosen to (1) allow full display of two 7.5' quadrangles at a time on the TV monitor (each of dimension 155 points by 200 lines), and to (2) give a map cell size of approximately 0.5 hectare (1.2A). These dimensions were then used to convert the map coordinates (in inches) to ground coordinates using the following formulas:

$$X_{nf} = \frac{X_m \times 155}{W} + (N \times 155) \tag{18}$$

$$Y_{nf} = \frac{Y_m \times 200}{L} + (M \times 200) \tag{19}$$

where

X_{nf} = X value in new file

X_m = X value on the map

Y_{nf} = Y value in new file

Y_m = Y value on the map

W = 7.5' map width in inches

L = 7.5' map length in inches

N = 0,1,2, or 3 depending on whether the 7.5' map is the first, second, third or fourth from the west side of the 30' block

M = 0,1,2,3 depending on whether the 7.5' map is the first, second, third or fourth map from the north side of the 30' block

The control point coordinates for the three Landsat dates and the new ground file were run through the regression program DANIEL. This program calculated the equations necessary to transform the Landsat data to the new ground coordinate file. These equations were of the form:

$$X_{Landsat} = b_o + b_1 X_G + b_2 X_G^2 + b_3 Y_G + b_4 Y_G^2 + b_5 X_G Y_G \tag{20}$$

and

$$Y_{Landsat} = b_6 + b_7 X_G + b_8 X_G^2 + b_9 X_G + b_{10} Y_G^2 + b_{11} X_G Y_G \tag{21}$$

where X_G and Y_G on the right side of the equation are new ground file coordinates.

The equations from DANIEL were used in the program COTRANS to place the Landsat data into the new file. This program resampled the data by using the DANIEL equations and the coordinates for each new file cell to predict the corresponding location in the original Landsat file. The data values for that pixel were then transferred to the cell in the new file. This was done for the 7/5 ratio bands for each date, the end product being a file 620 points by 800 lines with three bands: May, August, and October (when a 30' block was not covered by a single Landsat frame, the appropriate portions of each frame were resampled and then sewn together.) The data had been rotated, so that the new file corresponded to the map, with a north-south orientation, and transformed so that a particular cell represented the same point on the ground for all three dates.

Least Squares Analysis of Landsat to Map Registration

Registration of multiple Landsat dates to a north-south ground coordinate system was accomplished using control points as described above. A major question associated with this technique was how many control points were required to give a satisfactory registration. Evaluation of this problem based on repeated registration of the same area using differing numbers of control points would have required an impractical amount of analyst and computer time. As an alternative, regression equations of the same form as equations 20 and 21 were developed for differing numbers of control points. These equations were then used directly to estimate average registration accuracy for the set of all pixels in the map product. This analysis was applied on two block sizes, 30' and 1°.

The 30' Sutter Block (Figure 5-2c) was selected as the initial test area. This region represented a typical Central Valley agricultural/nonagricultural land use mix. Seventy-seven control points were selected over the Sutter Block in as uniform a pattern as possible, averaging approximately four to five control points per 7-1/2 quadrangle. After recording and verifying the (X,Y) ground and corresponding Landsat coordinates, regression equations of the form given previously were fit to the 77 control points for each of the three 1976 Landsat dates (May, August, and October). These equations were then used to generate an 11 x 11 matrix of expected (X,Y) Landsat coordinate pairs systematically covering the 30' block.

Next, the number of control points was reduced to 61, then to 46, 33, 16 and finally 8 by culling points systematically from the original 77. In each case, culling was performed by removing one control point from each 7-1/2 quadrangle.* $X_{Landsat}$ and $Y_{Landsat}$ regression equations were fit to each set of points (66,46,33,16,8) for each of the three Landsat dates. The resulting equations were used to predict Landsat coordinate pairs for the same 11 x 11 matrix of systematically located map reference points used previously.

Registration error introduced by reducing the number of control points below 77 was computed for each coordinate pair in the 11 x 11 matrix by subtracting the predicted value (based on k=66, 46, 33, 16, or 8 control points) for X or Y from its expected value (k_{max}= 77). That is

$$\Delta X_{ij(k)} = X_{ij(77)} - X_{ij(k)} \quad (22)$$

$$\Delta Y_{ij(k)} = Y_{ij(77)} - Y_{ij(k)} \quad (23)$$

where i and j represent the row and column indices in the 11 x 11 matrix of points systematically covering the area to be registered.

*A point was removed unless removing that point would leave no control point in that particular 7-1/2 minute quadrangle.

For the 8 point case, one point was taken from every other 7-1/2 quadrangle in a checker-board fashion.

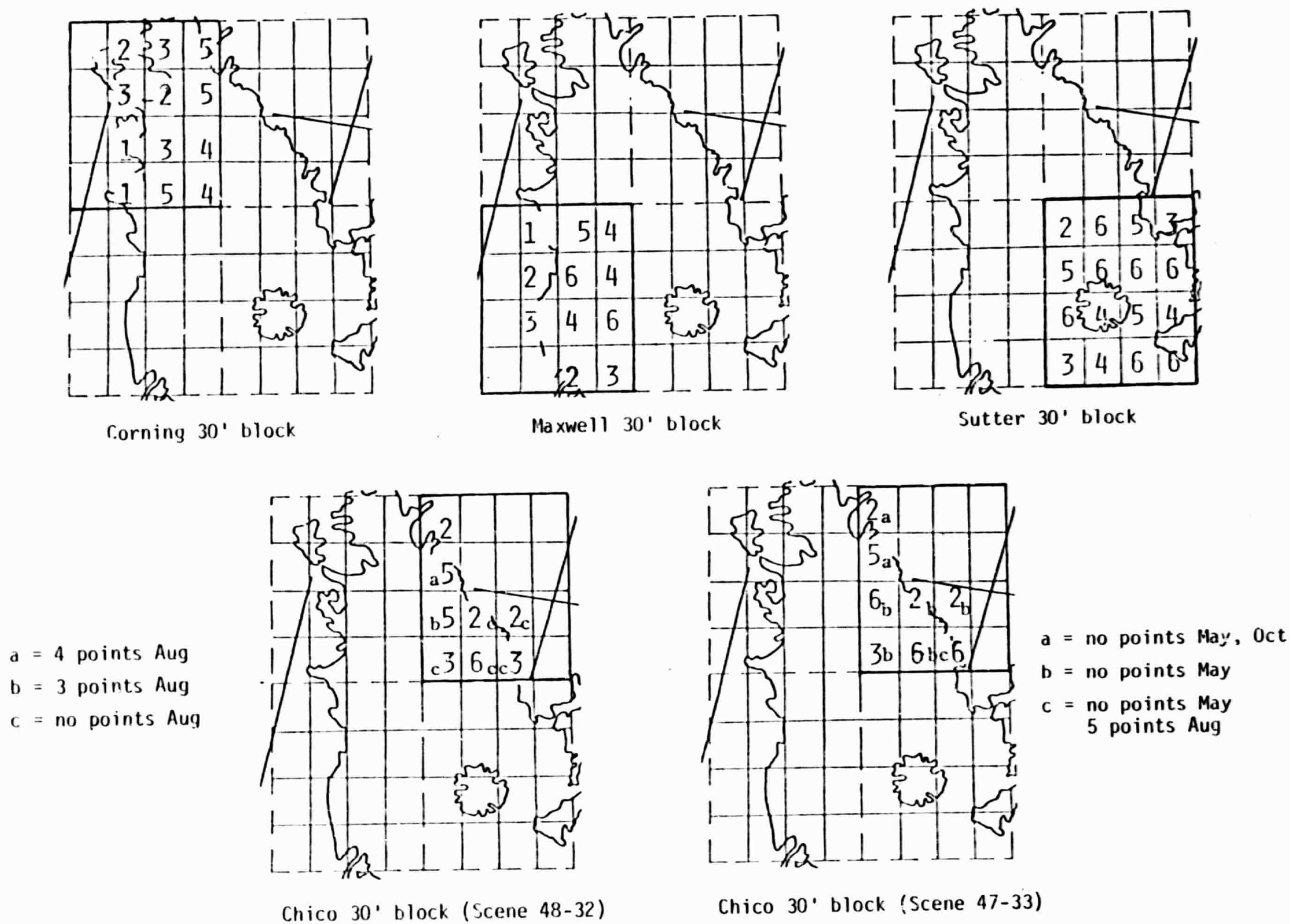


Figure 5-2. Number of control points for the 7.5' quadrangles within each 30' block.

The average registration error per pixel, sign ignored, introduced by $k < 77$ control points was then estimated by averaging the squared deviations and taking the square root, viz

$$\bar{d}_{X_k} = \sqrt{\left(\sum_{i=1}^{11} \sum_{j=1}^{11} (\Delta X_{ij(k)})^2\right) \div (11 \cdot 11)} \quad \text{and} \tag{24}$$

$$\bar{d}_{Y_k} = \sqrt{\left(\sum_{i=1}^{11} \sum_{j=1}^{11} (\Delta Y_{ij(k)})^2\right) \div (11 \cdot 11)} \tag{25}$$

$\bar{d}_{X_k}$ and $\bar{d}_{Y_k}$ were defined as the <u>relative</u> registration error for control point density k in the X and Y dimensions. That is, this component of error was defined in terms of differences from (relative to) Landsat coordinate positions predicted with $k_{max} = 77$ control points.

It should be noted that straight averages of $\Delta X_{ij(k)}$ and $\Delta Y_{ij(k)}$ will tend to give smaller values than $\bar{d}_{X_k}$ or $\bar{d}_{Y_k}$ due to cancellation of differences over the population of pixels to be registered. However, such straight averages are misleading. Classification error depends (in part) on the absolute misregistration for any pixel in the scene, not on an average, sign-considered registration error computed over all pixels.

The average <u>total</u> absolute error per pixel was defined as the Euclidean sum of the relative error ($\bar{d}_X$ or $\bar{d}_Y$) and a term representing the error associated with the regression model used to predict Landsat coordinate positions with 77 control points. This second error component was taken to be equal to the mean squared error (MSE) computed for these regression equations. Thus, assuming $\bar{d}$ and MSE are independent, the average total error per pixel was computed as

$$\bar{D}_{X_k} = \sqrt{(\bar{d}_{X_k})^2 + MSE_{X_{77}}} \tag{26}$$

$$\bar{D}_{Y_k} = \sqrt{(\bar{d}_{Y_k})^2 + MSE_{Y_{77}}} \tag{27}$$

for the X and Y dimensions, respectively.* In words,

$$\bar{D} = \text{Ave Absolute Error Per Pixel} = \sqrt{\begin{pmatrix}\text{error (bias)}\\ \text{introduced with}\\ \text{k points instead}\\ \text{of } k_{max}\end{pmatrix}^2 + \begin{pmatrix}\text{average of the square of the}\\ \text{deviations between pre-}\\ \text{dicted coordinates and actual}\\ \text{coordinates based on } k_{max}\\ \text{control points}\end{pmatrix}}$$

$\bar{D}_{X_k}$ and $\bar{D}_{Y_k}$ were computed for k=77, 66, 46, 33, 16, and 8, in the 30' Sutter block. The results for X are plotted in Figure 5-3a and for Y in Figure 5-3b. In addition,

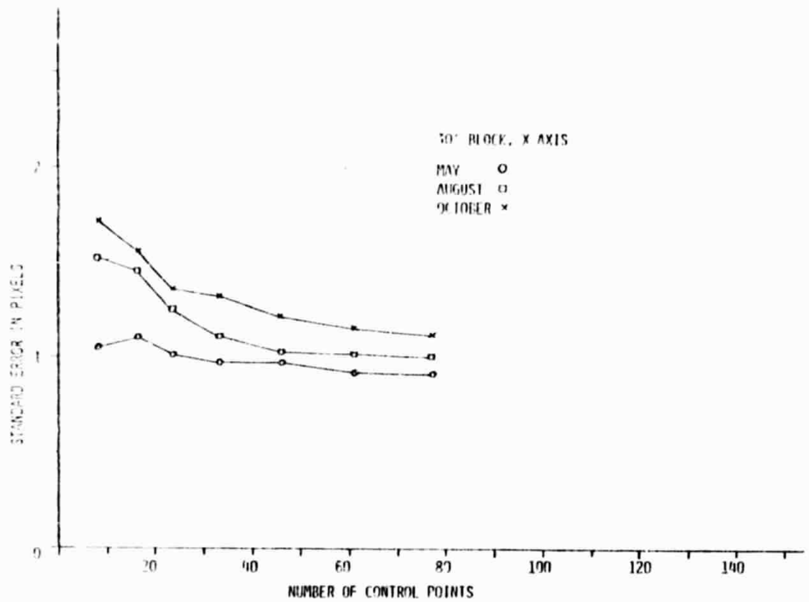
$$\bar{D}_k = \sqrt{\left(\frac{57\text{ m}}{79\text{ m}}\right)^2 \left(\bar{D}_{X_k}\right)^2 + \left(\bar{D}_{Y_k}\right)^2}\,, \tag{28}$$

the Euclidean average of the two errors (expressed in units of <u>vertical</u> pixels) was calculated and plotted in Figure 5-3c. Inspection of these figures indicated that registration error generally began to increase significantly below approximately 40 control points. Adding five control points to this number to account for culling of control points giving significant regression outliers, lead to a recommendation that at least 45 uniformly distributed control points be obtained for registration <u>on a 30' block basis</u>.

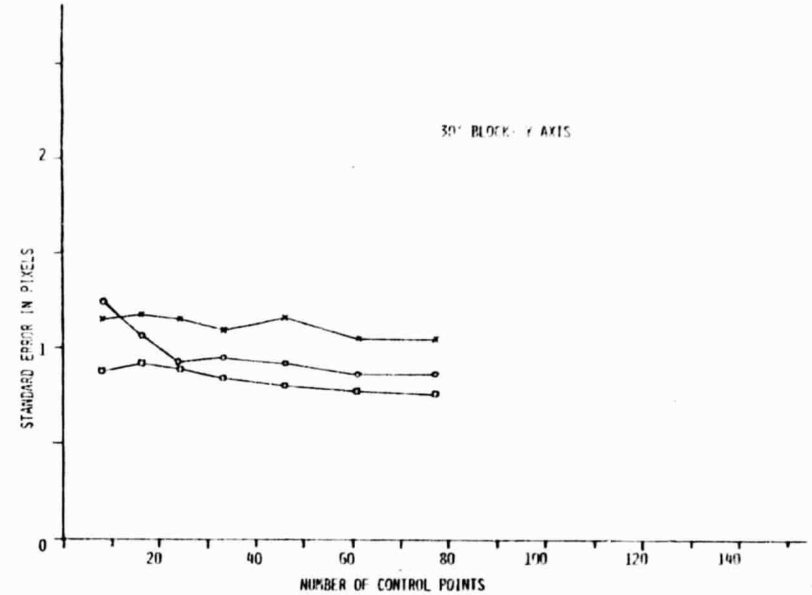
Three more 30' blocks were processed using a 45 control point objective. These were the Maxwell block (40 points after culling), the Corning block (37 points), and the Chico block (28 points from scene 48-32 and 32 points from scene 47-33) as shown in Figure 5-2a-e.

Visual inspection of the Sutter block registered with 77 control points showed that date-to-date registration error did not exceed one pixel. When this error occurred it was typically located along field boundaries originally flush with Landsat line-column geometry that had become diagonal in the new North-South coordinate system. The same situation obtained in the other three 30' blocks registered using the 45 point rule. As expected, wildland areas having very few control points tended to have larger registration errors.

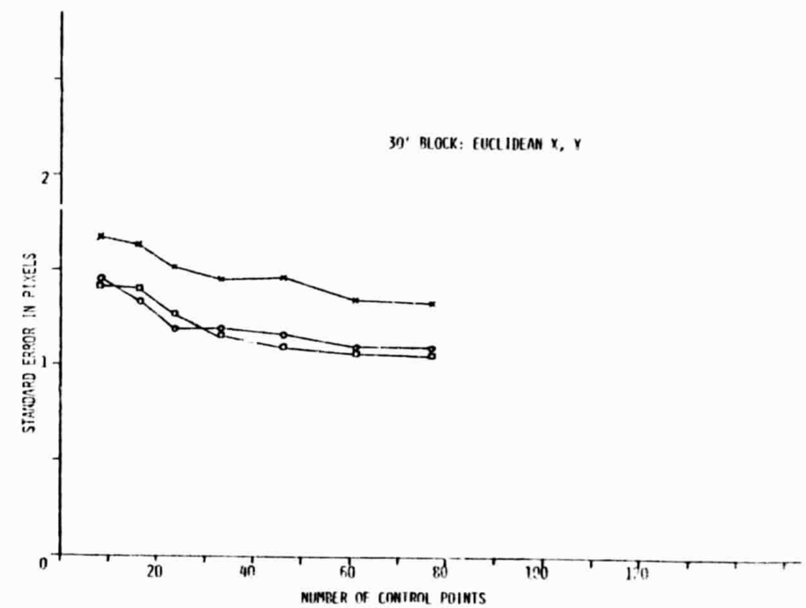
*A third component of error exists. This is the error introduced by predicting coordinate values away from control points with the regressions based on k_{max} = 77 points.



a 30' block, X axis



b 30' block, Y axis



c 30' block, Euclidean X, Y

Figure 5-3. Registration accuracy as a function of number of control points for Suiter 30' block.

Visual inspection of the four 30' blocks also indicated that Landsat-to-ground registration error was within error bounds predicted by $\bar{D}_k$. Field and road geometry in the North-South coordinate system appeared excellent. Seams between irrigated-nonirrigated class maps for the four 30' blocks were difficult if not impossible to detect, further confirming proper ground registration.

A further study was performed to determine if a less dense (and therefore less costly) network of control points would be required if registration was performed on an area larger than a 30' block. Time and resources permitted an examination of this problem only on the 1° block area (shown in Figure 5-1) used for the four 30' blocks registered earlier. Since this area was partially covered by two Landsat frames, and since separate registration regression equations were required for each frame, analysis of the large area registration problem was limited to the southeastern two thirds of the 1° block covered by scene 47-33.

The 7.5 minute quadrangles used in this registration problem are shown in Figure 5-4. Numbers inside each rectangle represent the number of control points available for registration within each quadrangle. Control points obtained in the previous work with 30' blocks were used to provide the best possible comparison of registration error.

A series of (x,y) regression equations of the form specified earlier were then computed using successively larger sets of control points. Thus the first pair of regression equations (predicting $X_{Landsat}$ and $Y_{Landsat}$) were based on one control point selected at random from every other quadrangle in a checker board fashion. The next pair of regressions were based on one control point from each quadrangle, the next on two points from each quadrangle, and so on.

Using the pair of regression equations based on the maximum number of control points (k_{max} = 149) as a reference, $\bar{D}_{x_k}$, $\bar{D}_{y_k}$, and $\bar{D}_k$ were computed for all other equations based on $k < k_{max}$ number of control points. These results are plotted in Figure 5-5a - 5-5c.

Based on the results of this latter study, it appears that registration satisfactory for producing irrigated-nonirrigated class maps can be obtained on a one half 1° block basis using 45-60 control points. The behavior of the regression relationships exlored here also strongly suggest a similar number for registration on a 1° block basis. As in the case of the 30' blocks, these control points should be spread in as uniform a manner as possible over the block and in the area surrounding the block.

An accounting caution on areas as large as 1° is advised. Suppose the resulting digital image or class map for the 1° block is represented by a rectangle having x columns and y rows, each (x,y) map cell of equal size. Then, in California's latitudinal range, a map cell in the top row will represent an actual ground area approximately one percent smaller than a map cell in the bottom row of the block. This effect is due to the convergence of longitudinal lines at the North Pole. Consequently, in producing area estimates, a correction (scaling factor) must be introduced by row (or by group of rows) to standardize the ground area represented by each map cell.

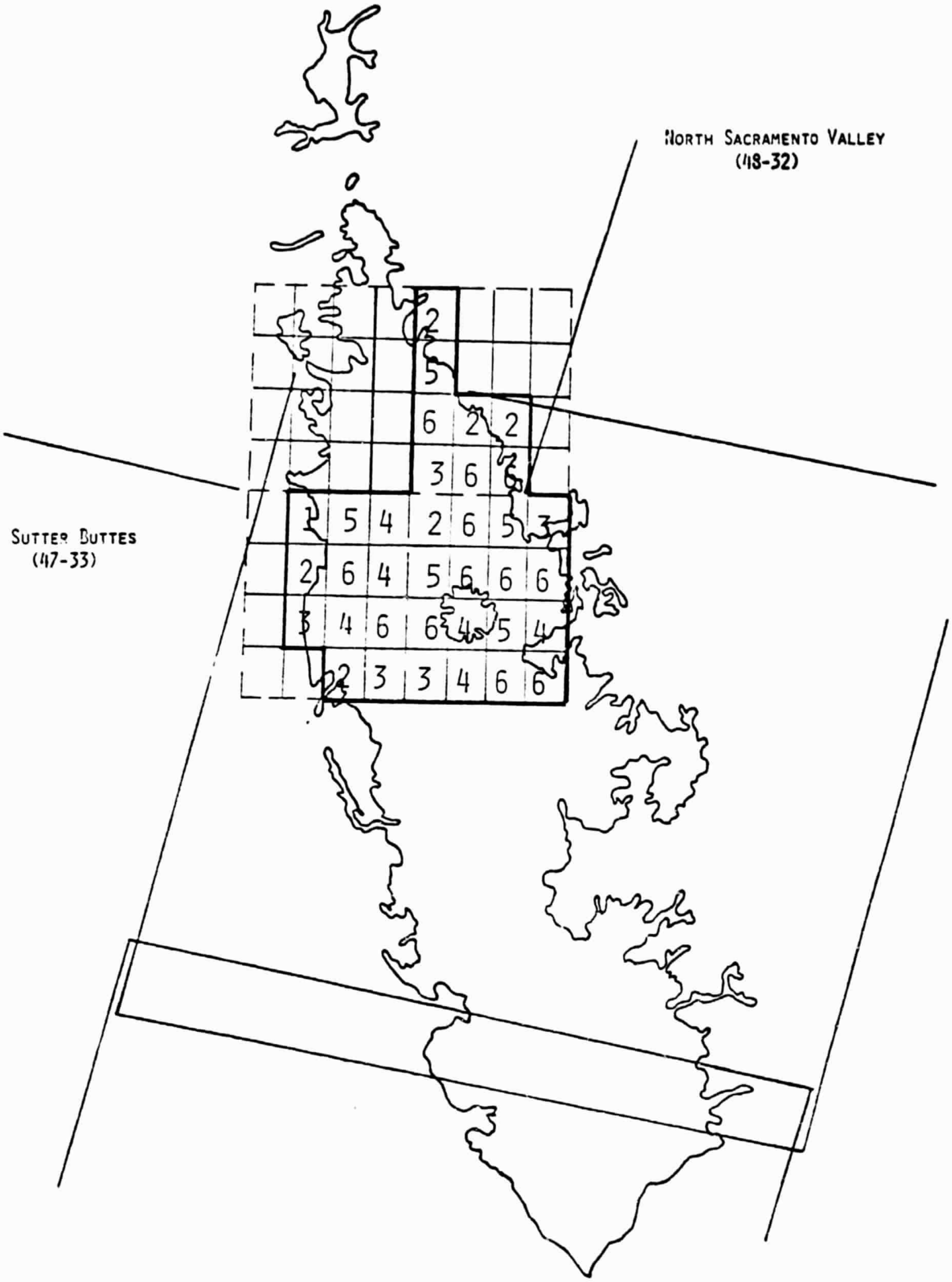
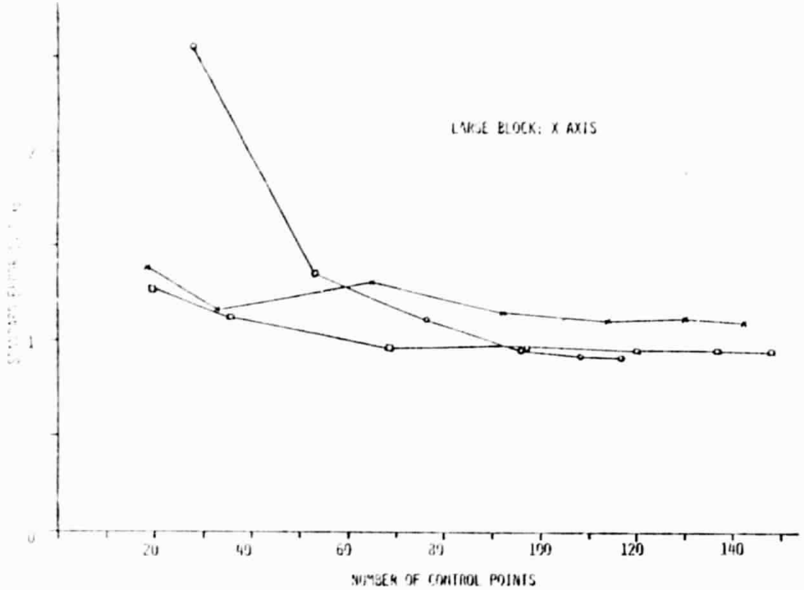
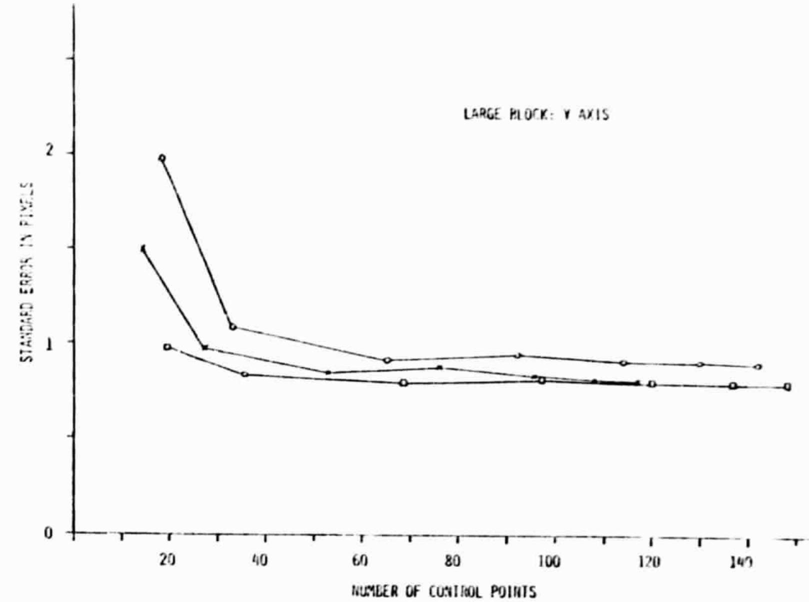


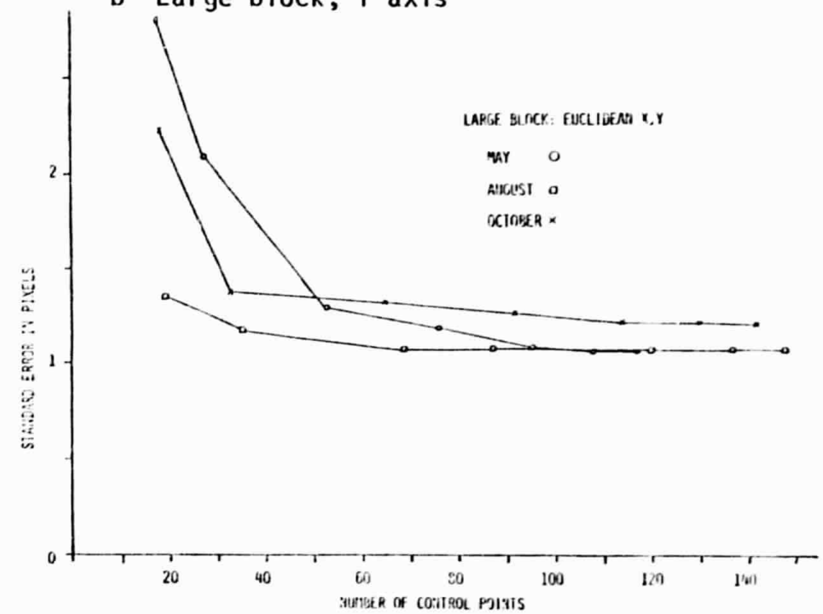
Figure 5-4. The number of control points within each 7.5' quadrangle available for least squares analysis.



a Large block, X axis



b Large block, Y axis



c Large block, Euclidean X, Y

Figure 5-5. Registration accuracy as a function of control points for large block covering parts of three 30' blocks.

5.1.2 Cross Correlation Registration Procedure-Geography Remote Sensing Unit (GRSU) U.C. Santa Barbara

The digital analysis of Landsat CCT's requires the ability to accurately register multidate imagery, particularly for identifying specific crops. As part of this year's effort, two registration procedures were explored — a manual procedure where tiepoints between two dates are selected by an analyst using a television monitor display and an automatic procedure using cross correlation and regression analysis.

Manual Approach

The manual procedure was undertaken on a portion of the statewide Landsat mosaic generated by the Jet Propulsion Laboratory (JPL) for the California Department of Forestry*. Using the 1° Sacramento West quad (August 1976) from JPL as a base, two other dates (30 May 1976 and 03 October 1976) were registered. Visual analysis of the three dates yielded 50 points that were observable on each date. The features were selected in a fairly systematic network over the entire scene.

Subimages from the base date, around each of the tiepoints, were then displayed on a video monitor and line and sample coordinates noted with a moveable cursor. The same subscenes from one of the other dates was displayed on a second monitor and the line and sample coordinates of the tiepoints noted. This procedure was repeated for the final date. When this had been completed for all 50 tiepoints, the appropriate line and sample coordinate values were used as input to GEOMA, a VICAR program designed to register digital images.

As a check, the 50 subimages for each date were overlayed and displayed on the video monitor after registration to determine the performance of the GEOMA procedure. Each date was displayed in a different color and alternately turned on and off. Those that were offset from the base date were re-evaluated to determine the correct tiepoint coordinates and the results placed in GEOMA for a second time. When the analyst had evaluated all 50 points and was satisfied with the fit of each, the three dates were considered registered and available for further analysis. While a detailed examination for goodness of fit was not conducted, the May and October image each seemed to be within 1 or 2 pixels of the base over the entire image. A greater number of tiepoints could improve this fit.

Cross Correlation Approach

Two tests were conducted using a VICAR cross-correlation procedure, PICREGB -- the first for a small area in Kern County and the second for a large area near Sacramento. The premise behind this approach is that stable features common to each date will drive the selection of accurate tiepoints. The presence of

---

*Managed through NASA-Ames Research Center as part of the California Integrated Remote Sensing System (CIRSS) Applications System Verification and Transfer Program.

noise in the images should not severely impact the procedure so long as the noise is random. Any adverse impact of noise can be minimized through image enhancement of stable features.

Such a procedure has been developed in LACIE for registration of multi-date agricultural scenes.[1] In most cases, the biomass content of fields changes dramatically from season to season. The major stable features are field boundaries, which can be accentuated by a digital high pass filter. In the LACIE approach, a high pass filter was applied to bands 5 and 7, and a binary image was created for each band with the lowest 85% of the pixels being set to 0 and the upper 15% being set to 255. The assumption here is that approximately 15% of the scene is made up of boundaries and edges. The edge images for bands 5 and 7 were added together to yield an image which showed the location of boundaries present on either or both bands. By using both bands, a greater population of edges was available for analysis. Composite edge images for multidates can then be registered using the cross correlation procedure.

Using a three - 7½ minute quadrangle study area in Kern County, a test of the cross correlation procedure was undertaken. Three dates of imagery (June, July, and October) were used. All three data sets had been previously registered, using visual techniques, to within ½ pixel accuracy so the test consisted of judging the ability of the cross correlation procedure to correctly select the corresponding tiepoint coordinates. Assuming a maximum misregistration of 0.5 pixels, the average expected misregistration in either the X or Y direction is 0.25 pixels, or up to 0.35 pixels in a diagonal direction. The 30 tiepoints were selected from over the scene in a systematic fashion. Each date was enhanced as described in the LACIE procedure with the exception that the high pass filter used enhanced only vertical features. The result was that north-south trending boundaries were emphasized while the east-west trending lines were not. While not specifically evaluated, the assumption that 15% of the scene consisted of edges was accepted.

The cross correlation was carried out on a window of 32X32 pixels on a base image window of 64X64 pixels around each of the 30 tiepoints. Those points for which a poor correlation was found were removed from the analysis. The remaining points, consisting of line and sample coordinates for the base date and the coordinates (output from PICREGB) for the date to be registered, were analyzed by a simple regression program that determines how well one group of points predicts the other. A residual of approximately 2 pixels was accepted as the upper threshold criterion for editing out bad tiepoints. After editing out those points with too large a residual value, the remaining points and their values from PICREGB were analyzed to determine the average Euclidean distance between the base date and the date to be registered. The results are shown in Table 5-1. For reasons not yet clear, the performance of the band 5 + 7 composite for October and July was not as good as that for band 5 alone.

---

[1] Grebowsky, G. J. 1978. LACIE Registration Processing: A Technical Description of the Large Area Crop Inventory Experiment (LACIE). LACIE Symposium Proceedings of Technical Sessions Volume 1. NASA Johnson Space Center. October 23-26, 1978: 87-97.

Table 5-1

Average Euclidean Distance (Pixels) Between

Tiepoints Selected By Cross Correlation

(Kern County 3-Quad Study Area)

Date 1	Date 2	Band(s)	Average Distance
June	October	5	2.37
June	October	5+7	0.93
June	July	5	0.91
June	July	5+7	0.77
October	July	5	0.60
October	July	5+7	0.65
June + October	July	5+7	0.48

The test was continued to determine the value of a multidate composite base image to which additional dates could be registered. By increasing the number of edges for registration, the performance of the procedure should improve. Table 5-1 shows the average Euclidean distance between a July (band 5 and band 7) edge image and a June and October (band 5 and band 7) composite edge image. As can be seen the apparent performance of the cross correlation procedure is improved due to a greater population of edges in the base image. Since the Boolean addition of two dates together is a relatively inexpensive procedure, a fine tuning of the automatic registration procedure can be done without incurring significant additional costs.

The second test of the cross correlation was conducted on a portion of the Sacramento 1° quadrangle generated by JPL. Using the JPL August 1976 scene as a base, a May and October date were registered. This test differed from the first in that it was over a larger area (1500 X 1000 pixels), the proportion of edges in the scenes was not set at an arbitrary value but was instead determined by visual examination of the high pass-filtered images and the procedure involved the registration of previously unregistered images.

Four tiepoints were selected, in general proximity to the corners of the images, in order to perform a rough registration. This was necessary to bring all three images into the same coordinate scheme. As in the Kern County test, highpass filter images were created for bands 5 and 7 for each date. A cutoff value was visually determined and the data was given a binary stretch. A composite image was formed for each date by boolean addition of bands 5 and 7 (Figure 5-6a-c).

Using a systematic grid of 126 tiepoints, spaced 100 pixels apart in the line and sample direction, the May and October images were analyzed to find those tiepoints that could be matched with the August base. PICREGB, a VICAR program, searches the area around the input tiepoint locations to find the best pixel-to-pixel match. If an adequate correlation is found for a particular tiepoint grid location, that tiepoint is removed during subsequent editing. Afterwards, those tiepoints with good correlation were analyzed by regression to remove any tiepoints whose spatial location was inconsistent with the overall pattern of tiepoints. The criterion developed was based on visual examination of the histogram of tiepoint residuals (observed-computed). In most cases, the peak of the histogram centered between -0.5 and 0.5. A cutoff was used when the number of tiepoints for a particular residual value fell below three and did not rise at least to a value of three within one residual unit (see Figure 5-7).

The remaining tiepoints were then used to register the May and October dates to the JPL base. Of the 126 tiepoints in the original systematic grid, there were 41 and 52 tiepoints retained for registration for May and October, respectively. Figure 5-8 shows a multidate color composite made from band 5 of each date.

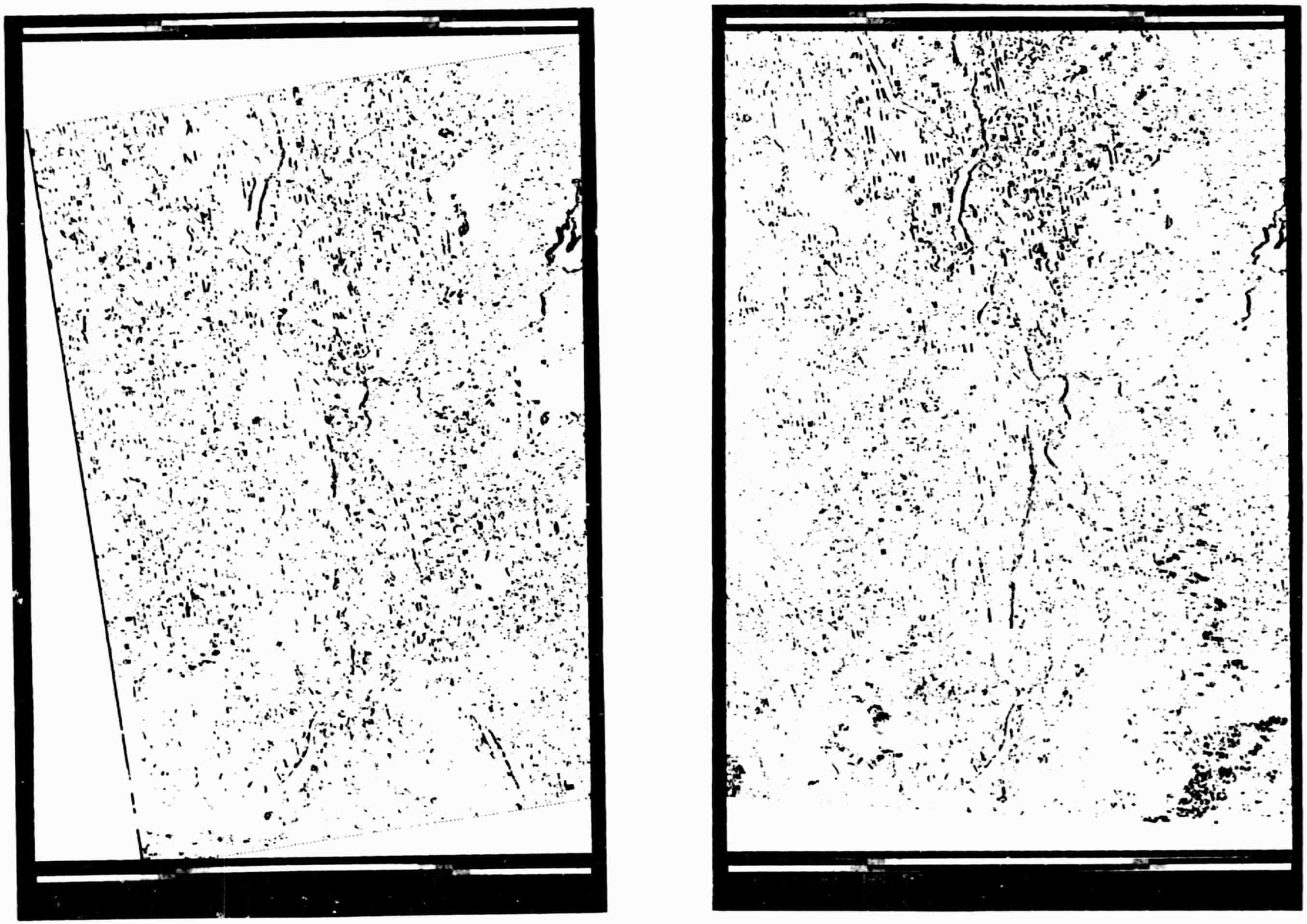
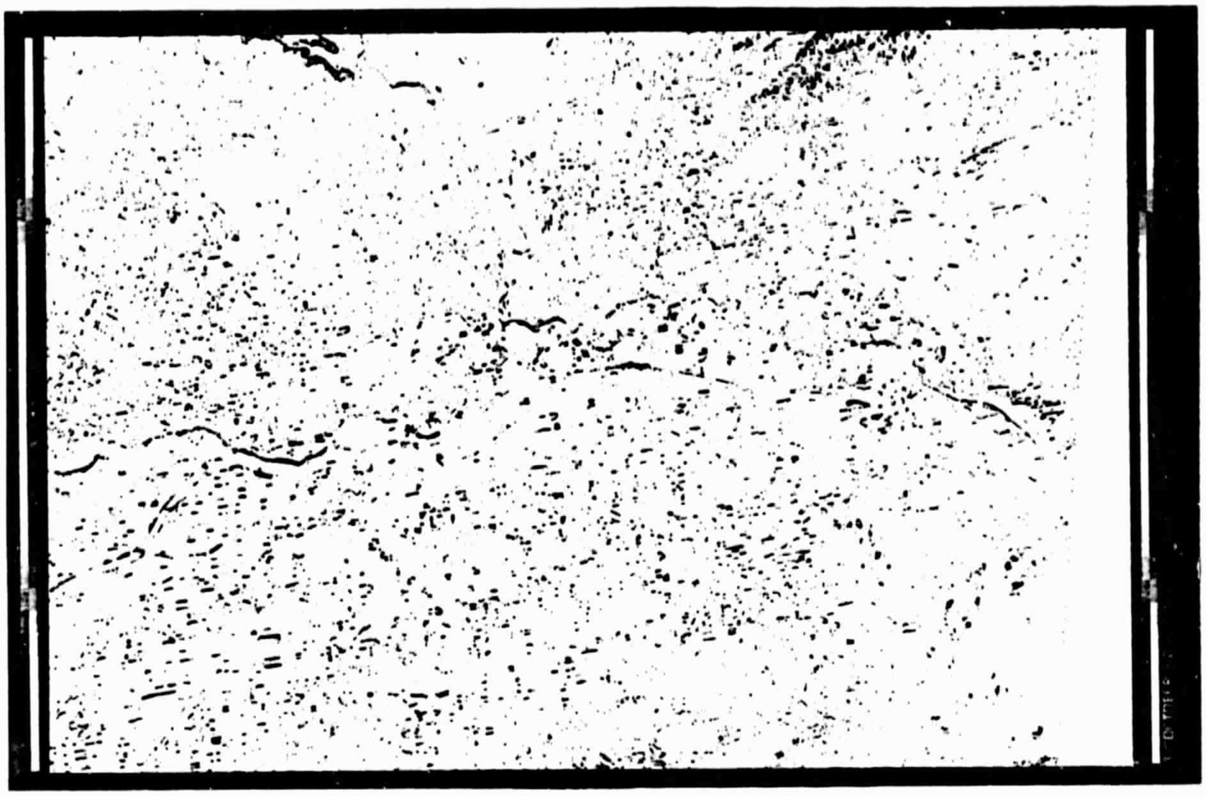


Figure 5-6 a b



c

Figure 5-6 (cont'd)

Figure 5-7.Histogram of Residuals for Line Coordinates
(October on August Base)

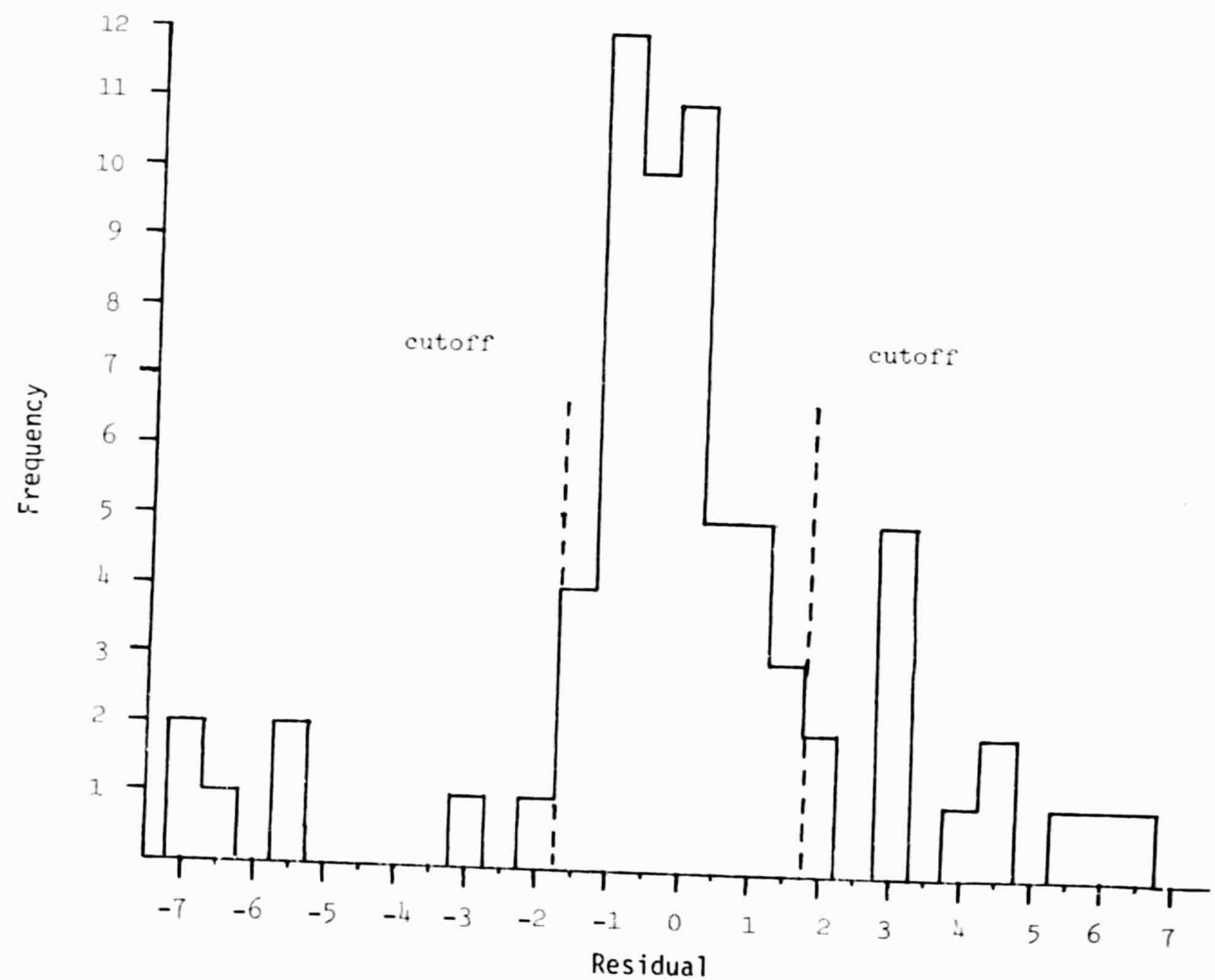


Figure 5-3

To test the goodness-of-fit between the registered images, each of the images (1500 X 1000 pixels) was divided into 24 subscenes (250 X 250 pixels). For each subscene, a feature that was identifiable on each date was located on the television monitor. Generally, this feature was a road intersection. Using a cursor, the line and sample coordinates of the feature were obtained on each date. Tables 5-2 and 5-3 show the coordinate values and the Euclidean distance between the May and August dates and the October and August dates. The average misregistration for May and October was 1.36 and 1.56, respectively. As can be seen, the majority of the points were within one pixel registration. The large misregistrations occurred in those areas near the edge, beyond the main body of tiepoints. This was seen particularly on the eastern boundary of the images where the cross correlation procedure, optimized for stable edges, did not yield tiepoints in the native vegetation of the Sierra foothills. For reasons not entirely clear, the procedure yielded few tiepoints on the northern boundary, although this area was covered by agricultural fields.

The computing cost for the procedure was approximately $100.00 (or $30.00 at the current overnight rate). This can be broken down for a 1500 X 1000 image as seen in Table 5-4.

Improvements in the procedure are possible. A greater saturation of tiepoints in the input grid would result in a greater number of tiepoints being retained. This can be done for the image as a whole or for selected problem areas (natural environments, image edges). A second improvement would involve localizing the cutoff procedure on the highpass filter image. For this work a global cutoff was used, but a moving block of 100 X 100 pixels, for example, would allow for local definition of edges. Presently, optimization of edge definition for one environment may be done at the expense of another.

A third possibility that remains to be examined is the use of non-binary input to PICREGB. While the current procedure used the binary approach proposed by LACIE, it may be that optimal edge definition lies somewhere between the binary and continuous extremes.

Improvements in the edge image inputs to PICREGB appear possible if a texture is used in place of one subjected to a high pass filter. Figures 5-9a and b show a texture and high pass image, respectively, for a 3-quadrangle area in Kern County. While a comparison between the two is limited since the high pass filter was operating in one dimension and the texture procedure is two-dimensional, the texture image has much less "sparkle" and its edges are more easily separable (in the spectral domain) from the non-edge background. The texture image is created by computing the standard deviation for a 3 x 3 kernel.

The present regression and residual analysis utilizes a global regression to detect tiepoints significantly different from the general pattern. A localized regression may result in a better measure of residuals.

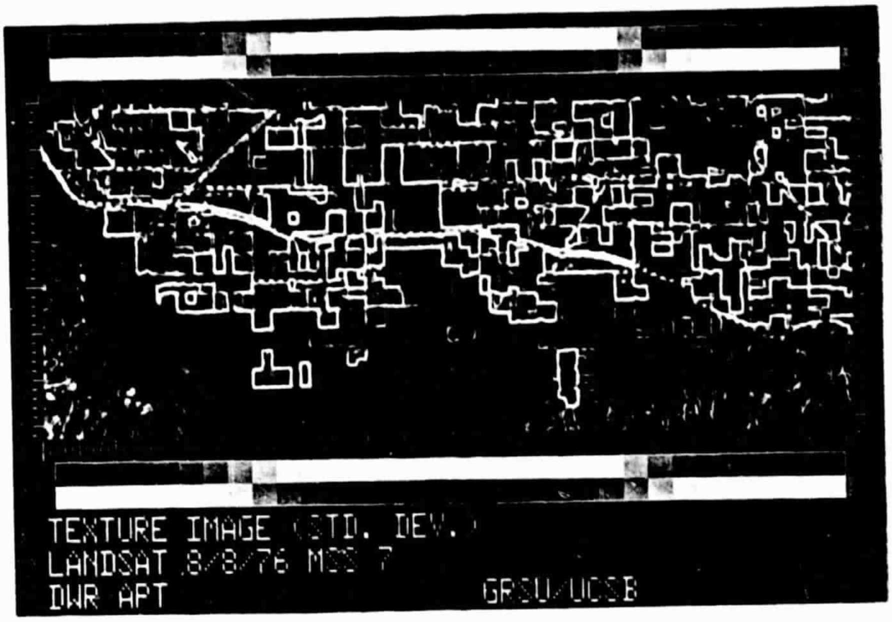
Finally, upon completion of this effort, it was discovered that TIECONP, a VICAR program that organizes the input tiepoints into vertices of triangular areas for localized "rubber-sheeting", broke down in those areas beyond the main body of tiepoint so that extrapolation of the computed fit was generally not valid toward the edges. A new version of this program, which does not have this extrapolation problem, has been received from JPL but was not used to re-run this analysis. The average pythagorean distance for those test points that were not beyond the perimeter of the main body of tiepoints was computed. The average misregistration from the August base was only 0.40 pixels and 0.54 pixels for May and October, respectively.

We feel that the procedure shows great promise as a cost effective technique for automated registration. The new EROS CCT registered format may preclude the need for an initial rough registration. A registration package could be developed that essentially automates the entire procedure.

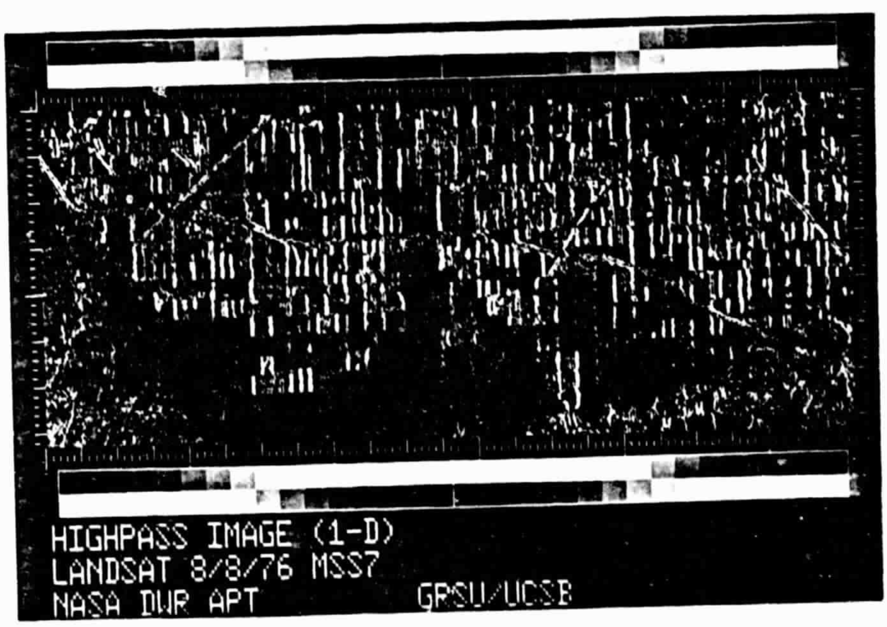
5.2 CLASSIFICATION OF MULTITEMPORAL LANDSAT DIGITAL DATA

The second major sub-task of Task II evaluated various classification techniques for estimating and mapping irrigated land within California. Three classification methods were analyzed this year to provide test results needed to recommend a system for a large scale demonstration (i.e., two hydrologic basins) in 1980. The algorithms evaluated were: (1) MSS band 7/MSS band 5 simple linear discriminant, (2) Kauth Thomas greeness transform, and (3) cluster labeling. UCSB studied all three techniques on a three 7.5' quadrangle test site in Kern County (southern San Joaquin Valley); UCB tested the 7/5 discriminant using the 1° block test site in the Sacramento Valley.

C-4


TEXTURE IMAGE (STD. DEV.)
LANDSAT 8/8/76 MSS 7
DWR APT
GRSU/UCSB

a


HIGHPASS IMAGE (1-D)
LANDSAT 8/8/76 MSS7
NASA DWR APT
GRSU/UCSB

b

Figure 5-9

Table 5-2

Cross Correlation of May to August Base

Registration Accuracy

(Sacramento Valley Study Area)

JPL Base		May		Euclidean
Line	Sample	Line	Sample	Distance
194	98	192	92	6.3
159	360	158	359	1.4
103	724	103	724	0
232	864	235	865	3.2
317	124	316	124	1
356	398	356	398	0
310	699	310	699	0
459	947	463	849	4.1
733	141	734	141	1
700	477	701	477	1
631	610	631	610	0
732	903	733	905	2.2
933	130	933	130	0
973	405	973	406	1
847	656	847	656	0
902	930	903	930	1
1203	187	1201	188	2.2
1038	312	1038	312	0
1100	694	1100	693	1
1128	839	1127	839	1
1330	182	1329	184	2.2
1381	357	1381	358	1
1432	611	1432	610	1
1296	834	1294	834	2

Table 5-3

Cross Correlation of October to August Base

Registration Accuracy

(Sacramento Valley Study Area)

JPL Base		October		Euclidean
Line	Sample	Line	Sample	Distance
194	98	194	95	3
159	360	162	361	3.2
103	724	105	714	10.2
232	864	235	864	3
317	124	316	123	1.4
356	398	355	398	1
310	699	309	700	1.4
459	947	463	949	4.5
733	141	733	141	0
700	477	701	477	1
631	610	631	611	1
732	903	732	903	0
933	130	934	130	1
973	405	973	406	1
847	656	847	656	0
902	930	902	929	1
1203	187	1203	187	0
1038	312	1038	312	0
1100	694	1100	694	0
1128	839	1128	839	0
1330	182	1329	183	1.4
1381	357	1380	357	1
1432	611	1431	610	1.4
1296	834	1295	834	1

Table 5-4

Cost of Computer Processing for Cross Correlation of 1500 X 1000 Pixel Area

Procedure	CPU (seconds)	Core-Weighted (K-Byte-Seconds)	Reader-Remote (Cards)	Printer-Remote (Lines)	Disk-2314 (I/O)	Disk-3350 (I/O)	Cost-Normal Rate ($)	Cost-Deferred Rate ($)
Rough Registration	70.73	50,520	34	44	858	238	25	7.50
High Pass Filtering	57.29	16,034	25	37	4	344	13	3.90
Cross Correlation (PICREGB)	52.92	82,890	286	1412	4	4272	39	11.70
Regression and Residual Analysis	3.44	3,393	101	551	103	239	2	0.60
Registration (GEOMA)	95.93	60,528	124	141	35	206	29	8.70

5.2.1 San Joaquin Valley Test Site - UCSB

Classification Using 7/5 Cutoff Approach

Our experience to date on Task I indicates that the Landsat identification of irrigated land in California is a relatively simple task because of the bright red color infrared appearance of crops against generally non-vegetated backgrounds. Task II calls for the digital implementation of this procedure. While standard classification procedures often make use of all four of Landsat's spectral bands, the cost of such an approach is significant when numerous dates are used and an area as large as California is involved. The use of simple discriminants of redness that do not depend on all four channels of spectral information,such as the 7/5 ratio, are an attractive means to significantly reduce the amount of data to be processed.

A test of a simple discriminant of "redness" (7/5 ratio) was undertaken to evaluate the ability of the discriminant to identify cropland on multidate imagery and to test the use of a Boolean scheme to summarize multidate results. The data selected for evaluation was a three 7½' quadrangle study area in Kern County. Three registered and rectified dates (6 June 1976, 21 July 1976, and 10 October 1976) were used. In addition, a crop map based on field data was available for the study site.

The 7/5 ratio is a particularly effective discriminant for irrigated cropland. Band 5 returns a comparatively low brightness signal for vegetation while Band 7 returns a comparatively higher brightness signal. The result of the ratio is a value that is considerably higher for healthy vegetation than for other classes. Figure 5-10 is a plot of Bands 5 versus 7 for 21 July, 1976. The distribution is very similar to that resulting from plotting Kauth brightness and greenness channels, shown in Figure 5-11. The relationship between the two vegetation indices is verified in Figure 5-12, which plots greeness versus ratio values and has a correlation of 0.91 (the plot shown is that of 5/7 ratio's, the inverse of 7/5 and thus inversely related to greenness).

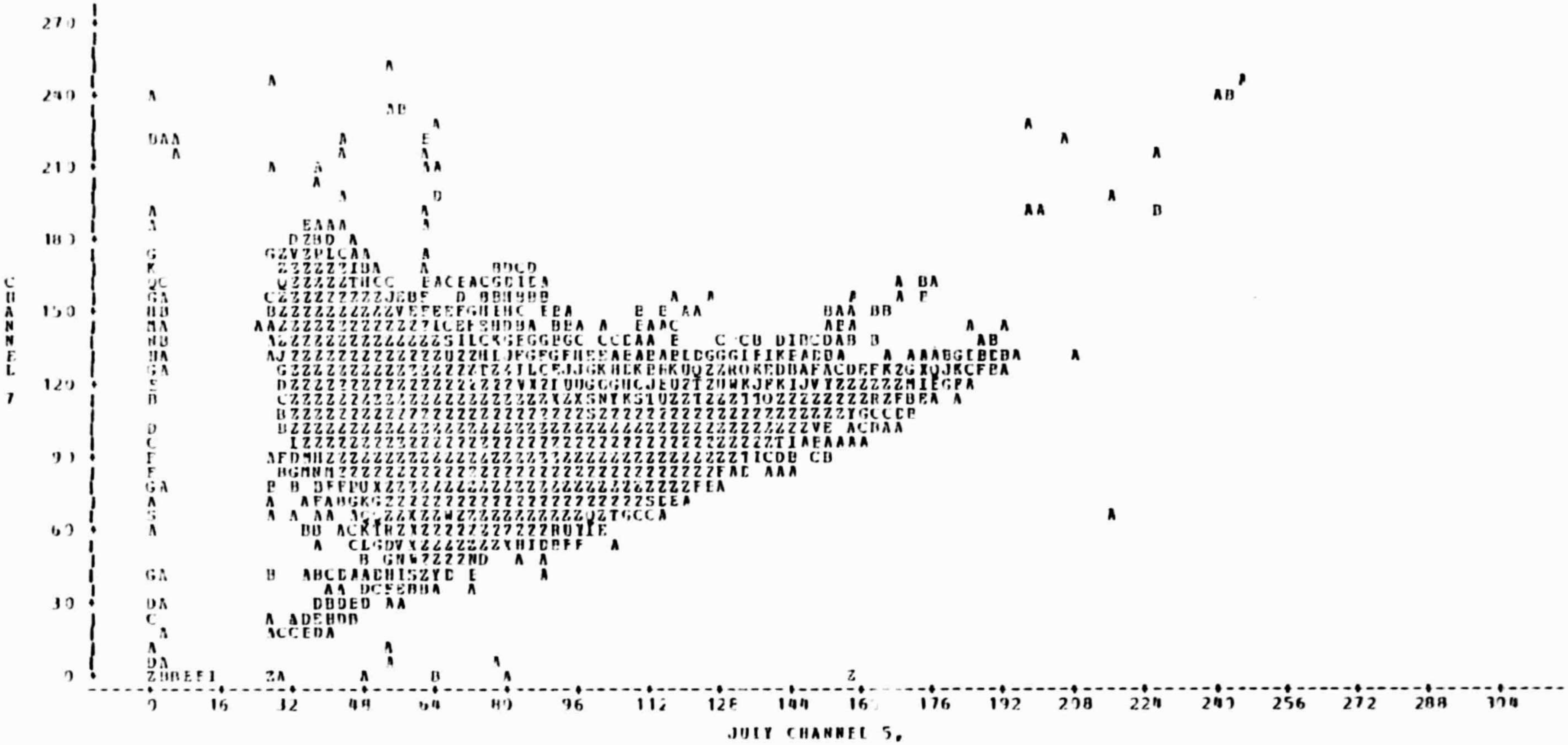
It is important to note that effective ratio cutoff values vary from date-to-date. In the spring, when there are numerous native grasses, the choice of the cutoff value must be conservative to avoid confusion. In the late summer or early fall, when most of the background vegetation is senescent, the selection of the minimally acceptable level of redness can be more liberal.

Using the selected ratio cutoff point for each date, classified images were created, with a value of 1 given to each pixel of irrigated vegetation and ∅ given to all others. In this fashion, images containing only vegetated cropland were created. Figure 5-13 shows such an image for July. The three registered dates were added together to result in a new image with four possible pixel values (0, 1, 2, 3) representing the number of dates on which healthy vegetation was found. When the three dates were added together, those areas for which irrigated vegetation was present on at least one date (sum >0) were flagged. When reduced to the binary case, this closely mimicks the decision process and type of final product from Task I.

It should be noted that each additional date added new information. The omission of any one date would have resulted in a smaller measurement of the amount of irrigated land. Because the earliest date used here was 6 June 1976, it is highly probable that a spring date would have increased the

WHEELER RIDGE DATA SET   JULY 21 1976
BAND 4 DNS VS. BAND 2 DNS

PLOT OF B7*B5   LEGEND: A = 1 OBS, B = 2 OBS, ETC.

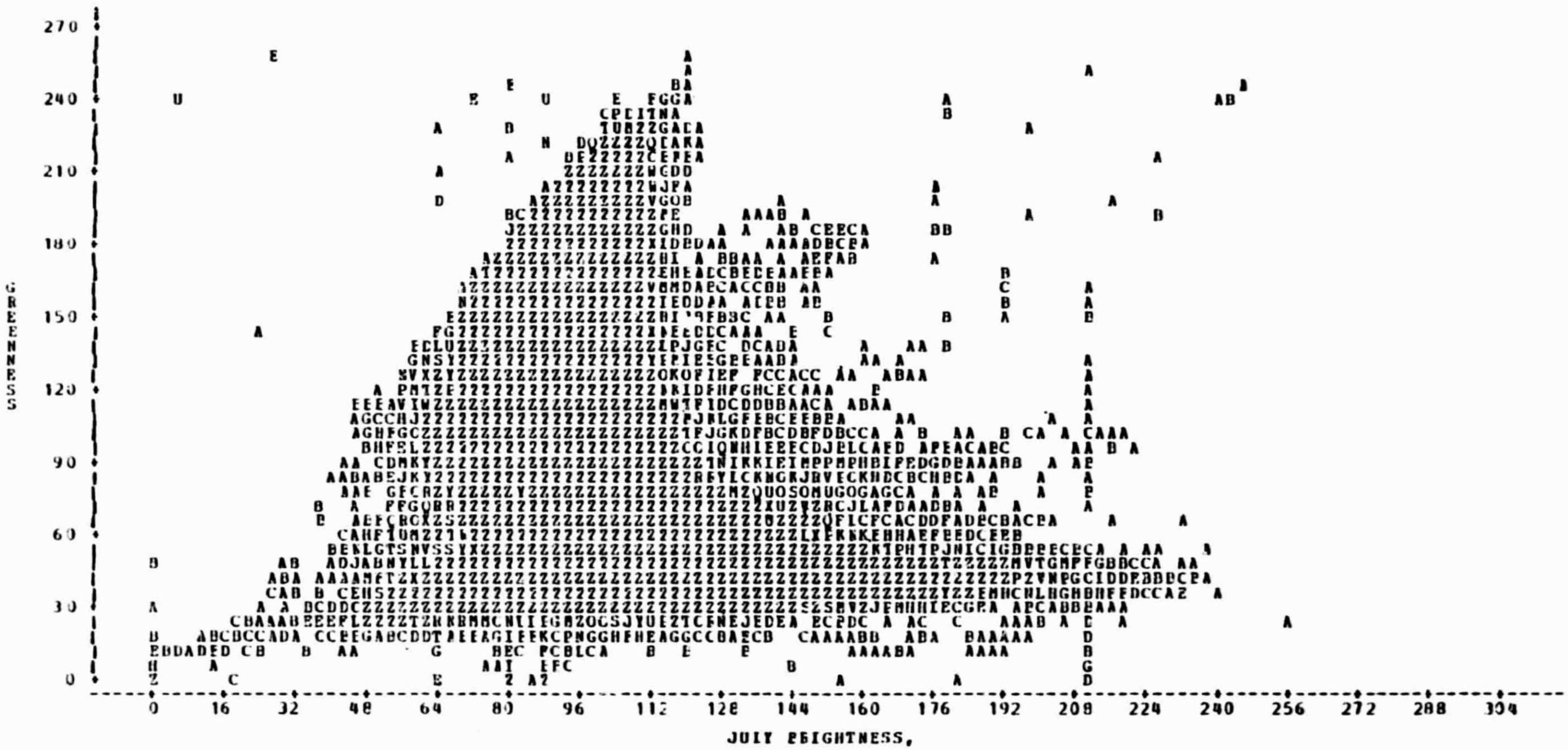


NOTE: 92834 OBS HIDDEN

Figure 5-10

WHEELER RIDGE DATA SET JULY 21 1976
GREENNESS VS. BRIGHTNESS

PLOT OF B7*B5 LEGEND: A = 1 OBS, B = 2 OBS, ETC.

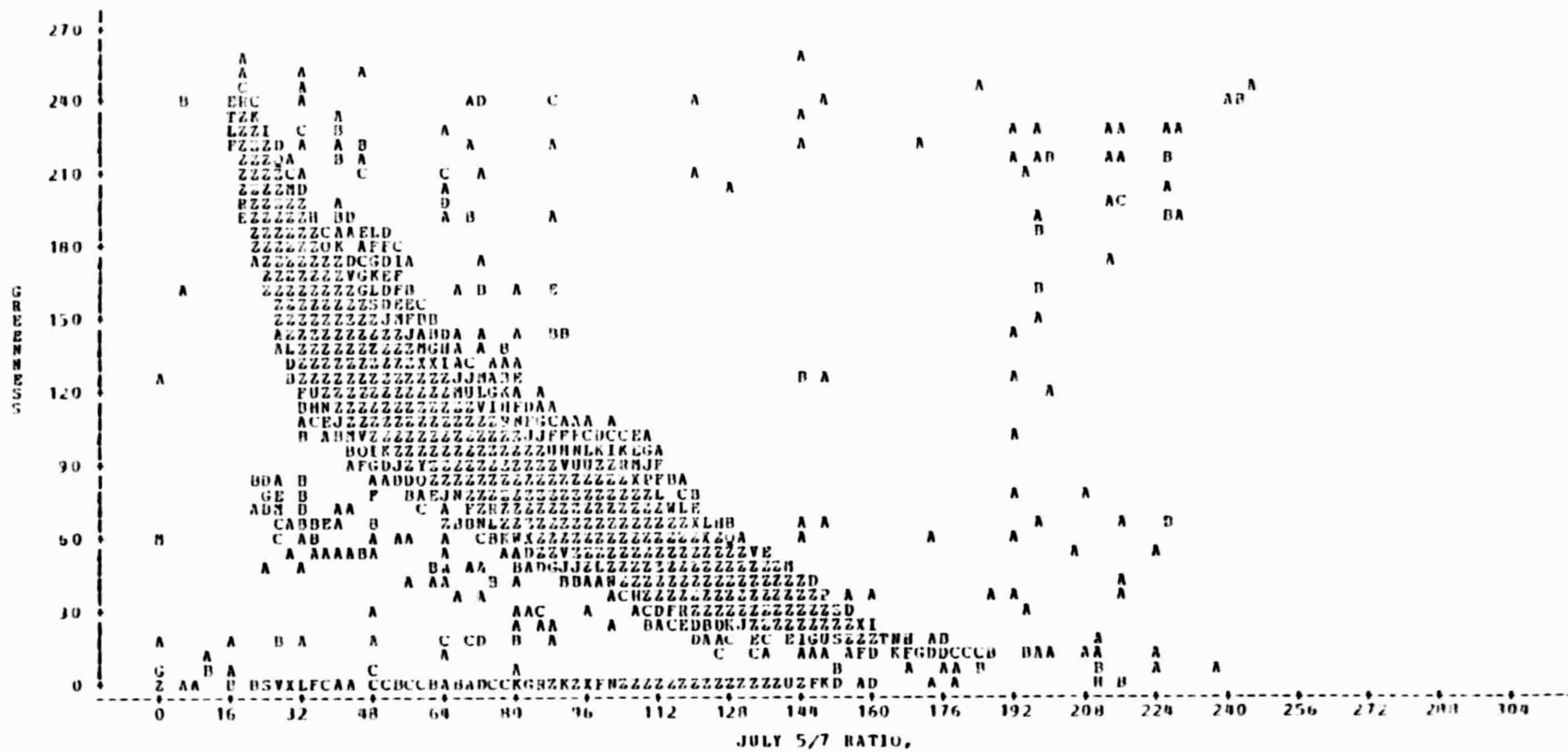


NOTE: 83502 OBS HIDDEN

Figure 5-11

WHEELER RIDGE DATA SET JULY 21 1976
5/7 RATIO VS. GREENNESS

PLOT OF B7*B5 LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 95076 OBS HIDDEN

Figure 5-12

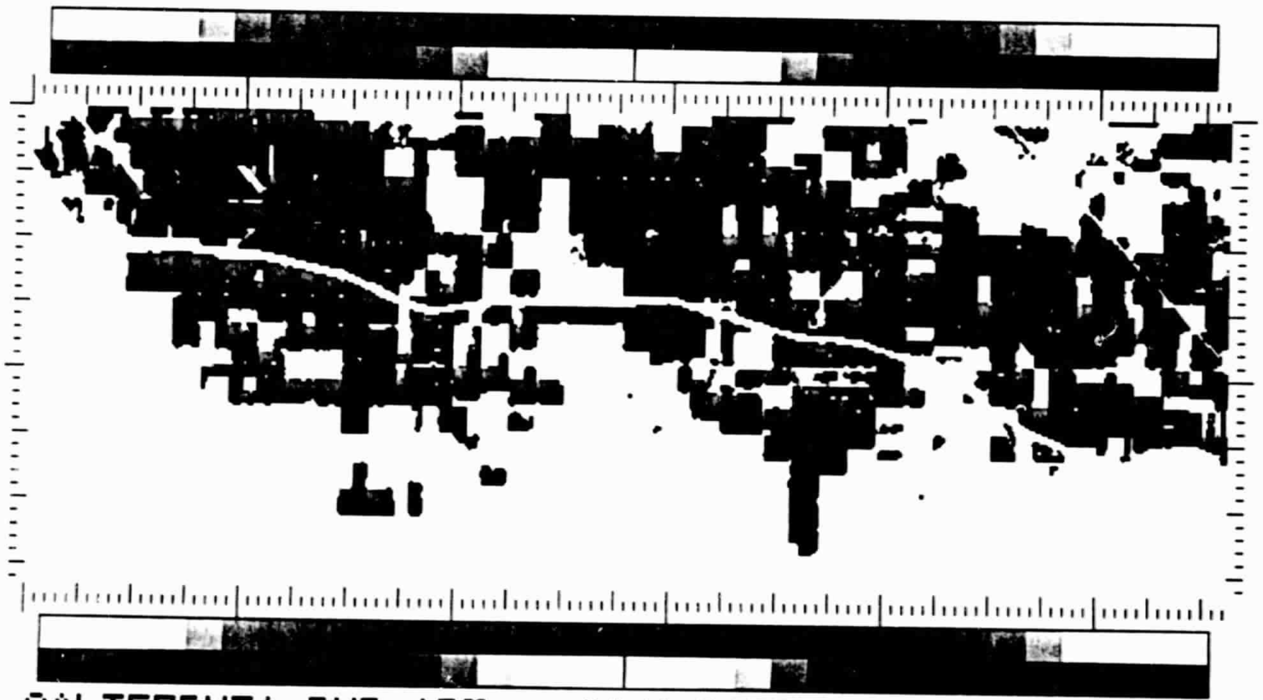


Figure 5-13

amount even more, because small grains were already senescent by June and were missed in this test of the 7/5 ratio.

Figure 5-14 shows those pixels that represent cropland irrigated on 0, 1, 2, or 3 dates. This information represents another level of sophistication in that the number of times a field was irrigated, and not merely the fact that it was irrigated at least once, is more valuable information for water demand determination. While not presented here, the technique could be carried a step further with classification using all possible permutations of dates. This would define a temporal signal of land use that possibly could be correlated with specific crop types much like the use of crop calendars in photo interpretation. Further work must be done in this area to determine the appropriate dates for such an effort. Whether reducing the problem to only the "greenness" and temporal domains is sufficient for crop identification must also be determined.

Our results indicated that, except for small grains, the ratio discriminant and Boolean classifier were adequate for detecting irrigated land. As mentioned previously, the use of a spring date would have caught the small grains (the available May date was not used at the time because of high gain settings in bands 4 and 5). Some apparent errors were noted where single pixels were classified as irrigated resulting in a slight salt and pepper effect. This would probably be accentuated on spring dates where isolated patches of native grasses are present. An editing procedure that removed clusters of only one or two pixels could be used to remove much of the salt and pepper pattern. This is of course what the manual interpreter does when using a minimum mapping unit.

The techniques demonstrated have the advantage of reducing the four channels of information from each date to a single channel. When multidate Landsat is considered, the number of possible band combinations also decreases. The reduced dimensionality of the data significantly decreases the cost associated with monitoring irrigated land. Because the study area was relatively small, there were no problems with the spatial extendability of the selected cutoff value.

The statistics for a 5 date analysis of the same area were computed using the 7/5 ratio and Boolean approach. In addition to the June, July and October dates used previously, images for 1 May 1976 and 8 August 1976 were added to the analysis.

Using Boolean addition, it was determined which dates yielded the best 1-, 2-, 3-, and 4-date estimate of irrigated lands (Figure 5-15). July proved to be the best single date for discriminating irrigated lands using the 7/5 ratio. A two date analysis increased the amount of irrigated acreage successfully discriminated by 12.6 percent with May and August being the best two dates. The addition of a third date increased the accuracy of the classifier by 4.3 percent. In this case, May, June and July proved to be the best combination although any combination of the May date with any two summer dates was close behind. Using four dates of Landsat - May, June, July and October — the accuracy was increased by an additional 2.8 percent. Obviously, the bracketing of the growing season with spring and fall dates is important to capture the temporal dynamics of cropping. The use of all five dates increased the classification accuracy to 97.4 percent when compared to a ground truth map. The addition of the fifth date improved the performance of the 7/5 ratio classifier by 1.8 percent.

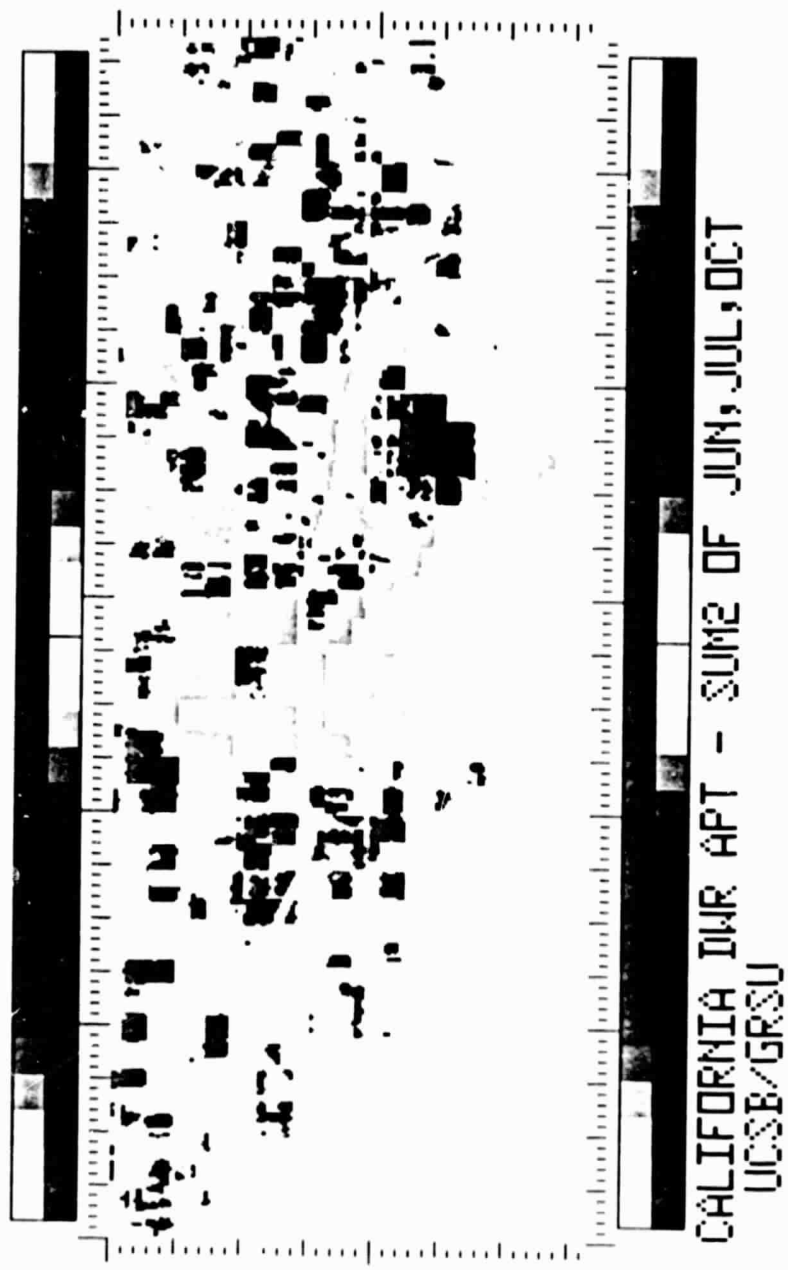


Figure 5-14

DATE SIGNIFICANCE TEST 7/5 DISCRIMINATOR
KERN COUNTY TEST SITE: ~ 50,000 ACRES TOTAL ~23,000 ACRES CROPLAND

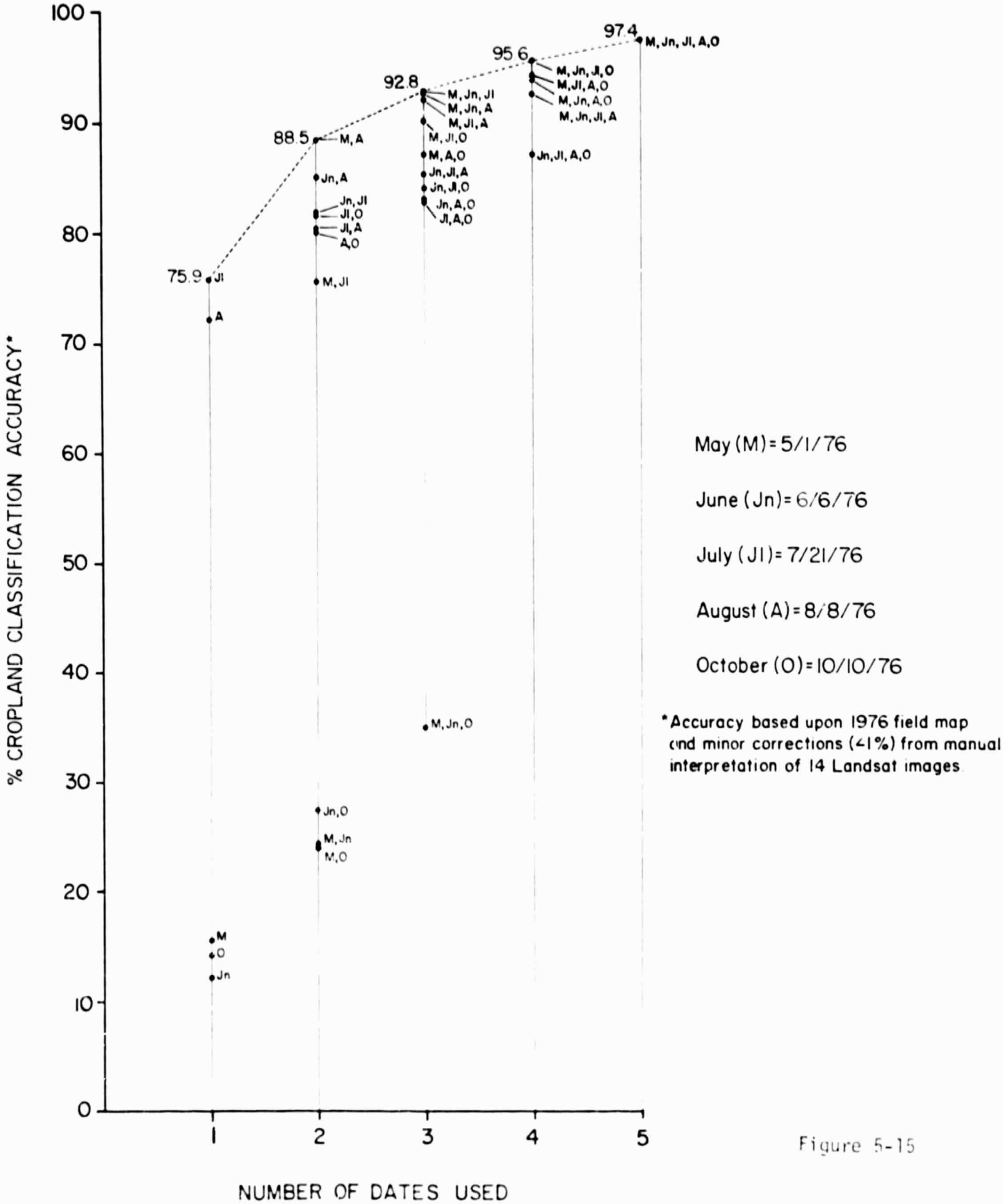


Figure 5-15

The ratio classifier appears to perform satisfactorily. The dimensionality of the decision process is greatly reduced by collapsing each date to a single channel of information. Because the decision logic requires the classification of a pixel as irrigated if its 7/5 ratio value is greater than a particular threshold value, it is important that an adequate number of dates be selected.

Classification Using Kauth-Thomas Greeness Transform

As shown in Figure 5-12, the 7/5 ratio and Kauth-Thomas "greeness" transform are highly correlated (r = 0.91). The greenness transform is a fixed linear transformation of all four Landsat channels and basically indicates a ratio of reflective infrared bands (MSS 6 and MSS 7) to visible bands (MSS 4 and MSS 5). Of interest to our research is whether this broader band ratio approach (i.e., greenness) is a more effective discriminator of cropland than the simple 7/5 ratio. Computational differences also need to be considered since the Kauth-Thomas transform requires substantially more processing.

We have created the Kauth-Thomas transform channels for all five dates of our Kern County data set. Using an interactive display program to determine optimum cutoff values for the greenness channel the three date analysis conducted for the 7/5 ratio has been repeated.

Figure 5-16 is the July image in binary form after the cutoff has been determined. This should be compared to Figure 5-13, shown earlier. Figure 5-17 is the sum of June, July, and October greenness classifications and should be compared to Figure 5-14.

The two sets of products visually compare very favorably, although a detailed statistical comparison with the ground truth has not yet been undertaken. When the processing of the digital ground truth map is completed, a systematic comparison of the two approaches will be initiated.

Classification Using Cluster Labeling

Because of the simple dichotomous decision by which cropland can usually be determined in most of California, our efforts have been oriented towards simple discriminants like the 7/5 ratio and greenness using a cutoff or threshholding approach to classification. The more conventional approaches to multispectral classification typically involve a maximum likelihood decision rule and/or the use of measurement space cluster labeling. Since these approaches provide a benchmark for comparison, we have also used them in our Kern County test site. Film products and a systematic comparison with the digital ground truth map will be generated during the next reporting period.

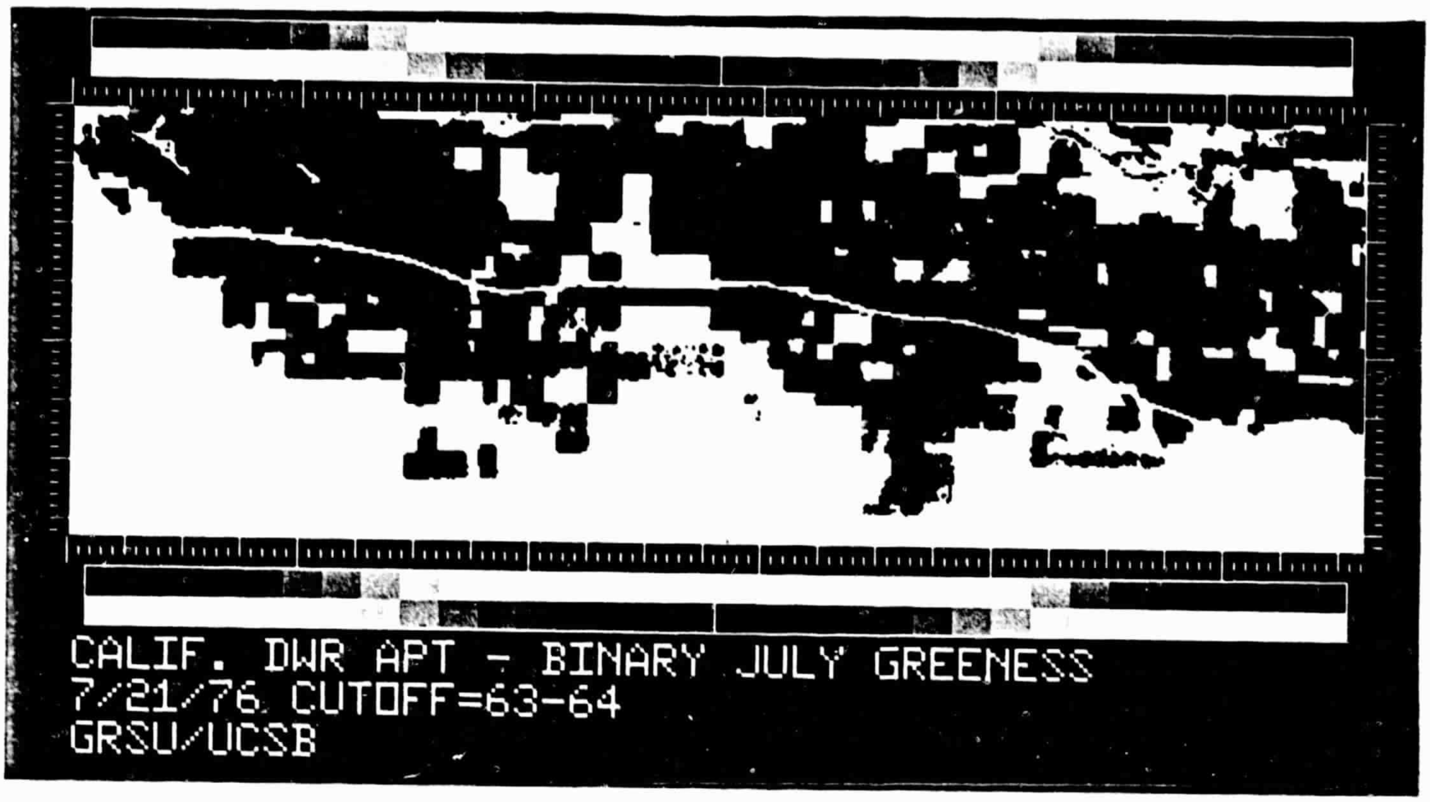

CALIF. DWR APT - BINARY JULY GREENESS
7/21/76 CUTOFF=63-64
GRSU/UCSB

Figure 5-16

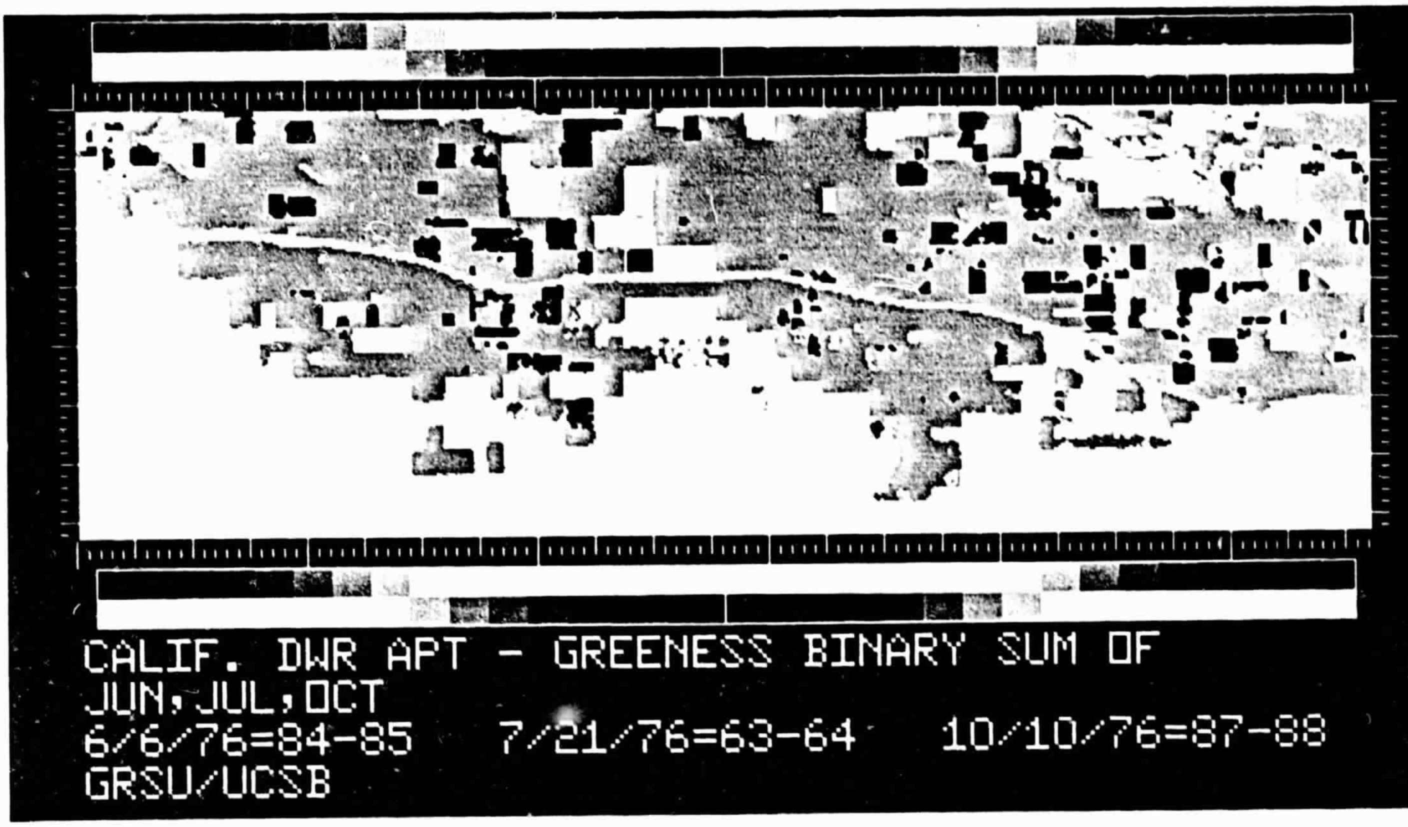

CALIF. DWR APT - GREENESS BINARY SUM OF
JUN,JUL,OCT
6/6/76=84-85 7/21/76=63-64 10/10/76=87-88
GRSU/UCSB

Figure 5-17

Summary of Developmental Efforts

In support of Task II activities, the following developmental efforts have been initiated or completed by the Santa Barbara group during this year:

- A sun angle correction procedure using a simple cosine function has been implemented in VICAR*
- A calibration procedure for matching Landsat I and Landsat II radiance values has been implemented in VICAR using values provided by ERIM.
- The ERIM XSTAR program which screens and corrects for atmospheric haze, has been reviewed for possible implementation in VICAR.
- Kauth transformations for "brightness," "greenness," and yellow stuff have been implemented in VICAR.
- Image to image cross-correlation procedures using existing VICAR programs are under review as one means for automating portions of the registration process.
- A more flexible, interactive environment, involving multiple video monitors, has been developed for rapid ground control point selection.
- Programs to interface our coordinate digitizer output to IBIS (Image Based Information System) format have been partially supported to assist our digitization of field data and conversion to image format in registration with Landsat data.
- A plotting program to view polygon data has been adapted to our facility (in support of digital field data conversion).
- An interface between VICAR formats (band interleaved by line and band sequential) and a statistical analysis package (SAS) has been written to facilitate conventional statistical examination of multispectral data.
- A program has been developed and implemented to allow interactive selection of a classification cutoff point, such as in the 7/5 ratio or greenness images.

Of a more general nature, a limited amount of support was also provided to assist the implementation of necessary driver and support programs for an optical-mechanical filmwriter. This device provides film output from digital data and is now being used quite extensively for this project.

---

*VICAR is the image processing package developed by JPL; it has been implemented on an Itel AS/6 at UCSB.

Evaluating classification accuracies for digital techniques requires the systematic comparison of field data, or "ground truth", with image classification results. By digitizing field boundaries and converting this data into a raster or grid format it is possible to overlay ground truth with Landsat data and systematically analyze the composite data set in a geobased information system structure. An interactive package of coordinate digitizing programs has been developed at UCSB to provide input to IBIS (Image Based Information System). Figure 5-18 is an example of a crop map for our three quad test site in Kern County after field borders have been digitized and rasterized. Subsequent processing assigns a class number to each field based upon field data. Once in the grid format this data can be processed by the full complement of VICAR and IBIS programs.

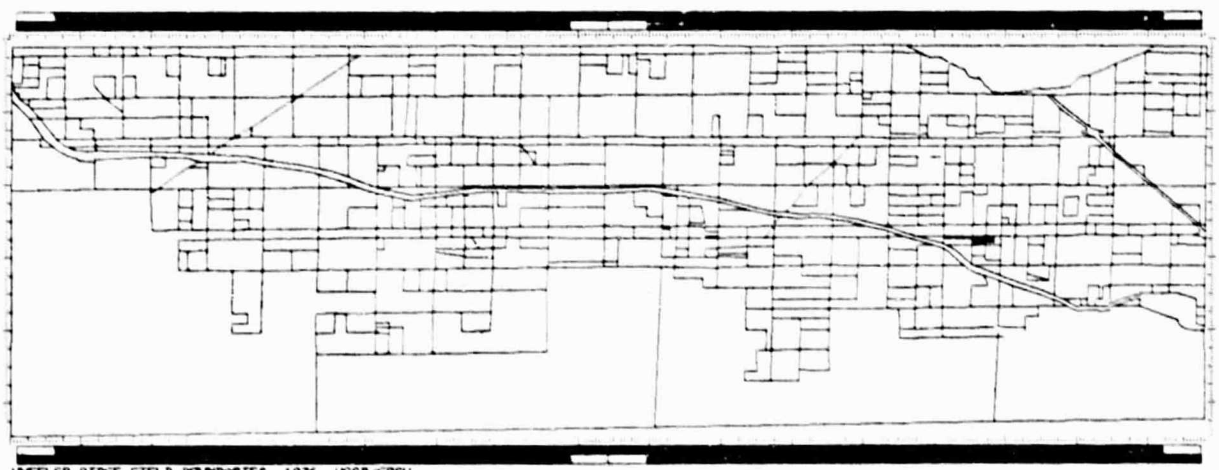


Figure 5-18

5.2.2 Sacramento Valley Test Site - UCB

The 1° registered multitemporal data set described previously (Section 5.1.1) was used for a relatively large scale test of the utility of the MSS 7/MSS 5 discriminator. The analysis was composed of three major parts: (1) creation of the classified 7/5 output; (2) analysis of the classification for use with a regression estimator; and (3) analysis of the mapping accuracy of the classified output.

MSS 7/MSS 5 Ratio Classification

As described in Section 5.1.1, the multitemporal data set used for this sub-task consisted of four files 620 points by 800 lines with three 7/5 ratioed bands; one band from each of three acquisition windows (May 22 and 30, August 28 and 29, October 3 and 4). The data had been rotated to a north-south orientation and transformed so that a particular cell represented the same point on the ground for all three time periods. Each 30' segment was made up of a 4 x 4 matrix of 7.5' quadrangle areas. Combining the four 30' segments (Chico, Corning, Maxwell and Sutter) produced the total 1° block.

By 30' block, the 7/5 ratio bands for each date were analyzed and a threshold value selected to separate irrigated from non-irrigated acreage. It was expected that the 7/5 threshold value would vary by date and ground location of the 30' block due to: (1) changes in the condition of annual grasslands bordering the area; (2) changes in type and proportion of crops grown; and (3) shifts in crop calendars due to climatic and latitudinal variations. Using the RSRP interactive image display system, each 30' block was displayed and analyzed separately.

To set the threshold value for a given band (date) the 7/5 data displayed on the TV monitor was compared to the DWR ground data (100% ground data was available for the entire area). Using a real time masking option, picture elements with values below a specified 7/5 value were masked out. This value was adjusted until the area shown as irrigated on the display corresponded as closely as possible to the irrigated area on the ground data maps. To further refine the threshold value selection, statistics (mean values and ranges of values) were obtained for the major crops, native vegetation and grassland in the area.

The threshold values for each date (Figure 5-19) were used to create an irrigation class map for each 30 minute block. For a given date, the 7/5 ratio of each pixel was compared to the selected threshold value and was labeled as irrigated if its value was greater than the threshold. After each pixel was labeled irrigated or not irrigated on all three dates, the bands were combined to create a class map. The three date pattern of irrigation for

each pixel was then labeled as one of 8 classes:

Class 1: not irrigated on any date

Class 2: irrigated in October only

Class 3: irrigated in August only

Class 4: irrigated in August and October

Class 5: irrigated in May only

Class 6: irrigated in May and October

Class 7: irrigated in May and August

Class 8: irrigated in May, August and October

The resulting class map was displayed and exclusion areas, such as wildlife refuges, wildland areas, and large urban areas, were masked out. These exclusion areas were determined using DWR's 7.5 minute quadrangle land use maps. The boundaries were transferred to the digital data using a real time polygon delineation routine controlled by the cursor. The exclusion areas were masked out and for display purposes relabeled as non-irrigated.

CORNING May = .98 Aug. = 1.27 Oct. = 1.24	CHICO May = 1.11 Aug. = 1.27 Oct. = 1.24
MAXWELL May = .98 Aug. = 1.27 Oct. = 1.24	SUTTER May = 1.21 Aug. = 1.27 Oct. = 1.24

7/5 VEGETATIVE INDICATOR

Figure 5-19.

A class map for the 1° block was created by sewing together the four 30 minute segments. An empty file, 1240 points by 1600 lines, was created and the four 30 minute blocks (each 620 points by 800 lines) were transferred to the proper location. The thresholded, 7/5 1° block was displayed in two ways. First, as a two class map (Figure 5-20a), irrigated or not irrigated, where classes 2 through 8 were combined into one class labeled irrigated. Second, where the 8 classes were differentiated, showing the temporal pattern of irrigation (Figure 5-20b).

The classification was then summarized by 7.5 minute quad to output a measurement of the proportion irrigated for each quad. Within each 7.5' quad, pixel counts were summarized for each of the 8 classes. Using the quad summaries, the accuracy of the 7/5 discriminant results when used with a regression estimator was assessed.

Accuracy of the Regression Estimator

For each of the sixty-four 7.5' quadrangles in the 1° block a measurement of the percent irrigated (Figure 5-21) as well as DWR's ground truth was available. Using the 7.5' quadrangles as sample units, it was possible to estimate the parameters of the regression estimator and its variance (Section 4.1.3, Equations 2 and 2a, respectively). The estimates were made both with and without stratification.* For comparison, estimates using 1 x 5 mile sample units were also made.

The analysis showed that good estimates (±5% @ 95%) can be achieved with as few as fifteen 7.5' quadrangles as compared to the fifty 1 x 5 mile SUs of Task I (Figure 5-22). However, these 15 quadrangles represent an area of approximately 233,107 ha (576,000 A) compared to the 64,752 ha (160,000 A) of the 50 Task I SUs needed to achieve ±5% @ 95%. An important part of the continuing work on Task II will be determining the appropriate size of SUs for digital analysis procedures.

Mapping Accuracy

One advantage of the regression estimator is that it corrects for bias (difference from truth) at the Landsat phase. Thus, if the percent irrigated is consistently over or under measured on Landsat, the regression estimator will give an unbiased estimate without an increase in variance. In generating an accurate <u>map</u> (as opposed to an accurate <u>estimate</u>), however, bias can be very detrimental. Map accuracy depends on <u>minimizing</u> miscalls: (1) errors of <u>omission</u> (missing land that was actually irrigated) and (2) errors of <u>commission</u> (classifying land as irrigated that actually was not).

---

*The stratified case used two strata based on the Landsat percent irrigated in each 7.5' quadrangle: stratum A had 60% irrigated and stratum B had less than 60% irrigated. Stratum A generally included agricultural practice strata 1, 2, 7 and excluded areas, while stratum B generally included strata 3, 4, 5, and 6 (Figure 4-7).

Black = Non-Irrigated

Red = Irrigated

Black = Non-Irrigated

Blue = Irrigated on May only

Pink = Irrigated on August only

Purple = Irrigated on October only

Green = Irrigated on May and August

Tan = Irrigated on May and October

Red = Irrigated on August and October

White = Irrigated on all three dates

Figure 5-20. Sacramento 1° Block showing land labeled as irrigated. The grid superimposed on the classified output outlines 7.5' quadrangles.

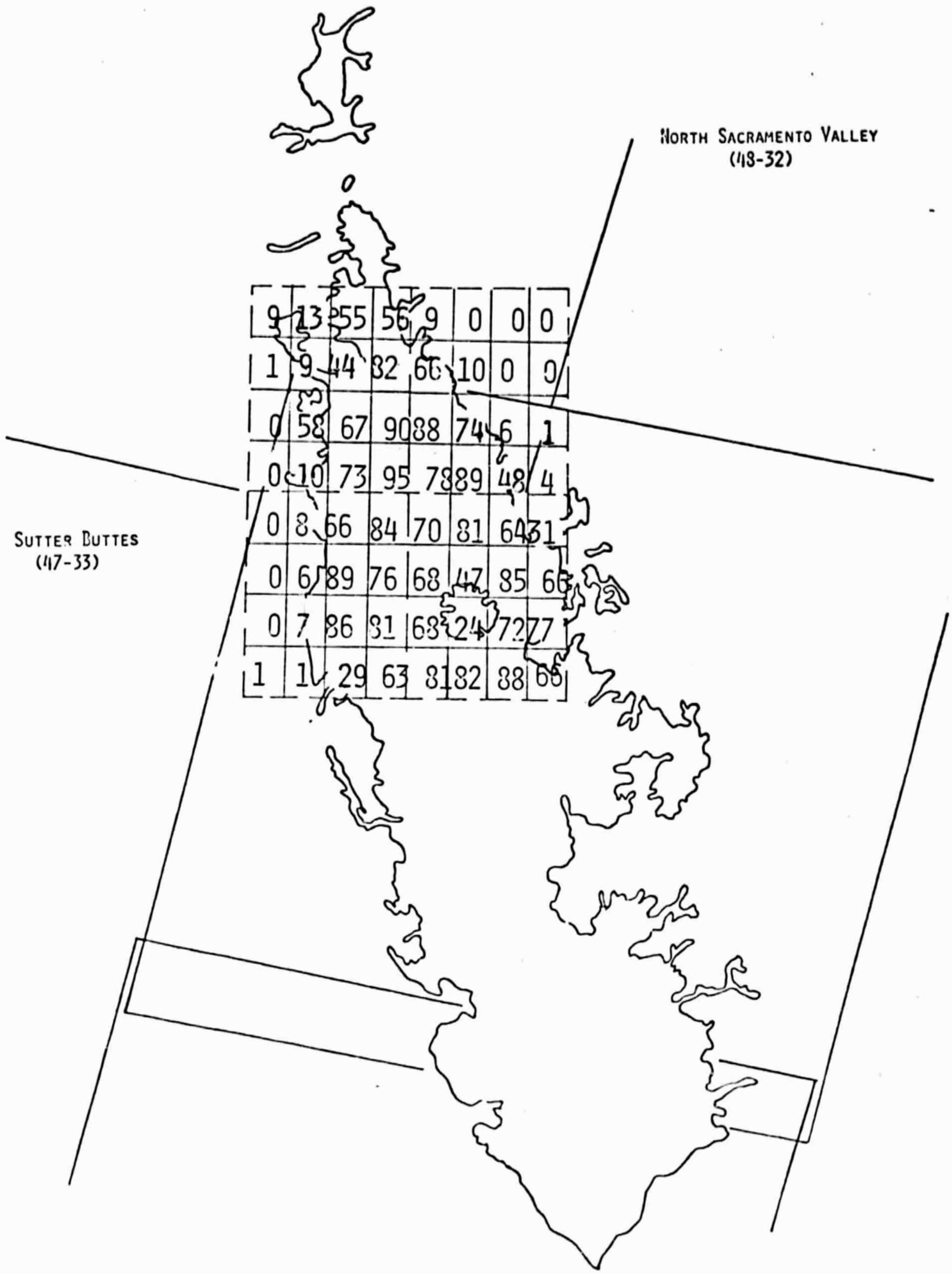


Figure 5-21. Percent irrigated based on classification of Landsat for each of the sixty-four 7.5' quadrangles in the 1° test site.

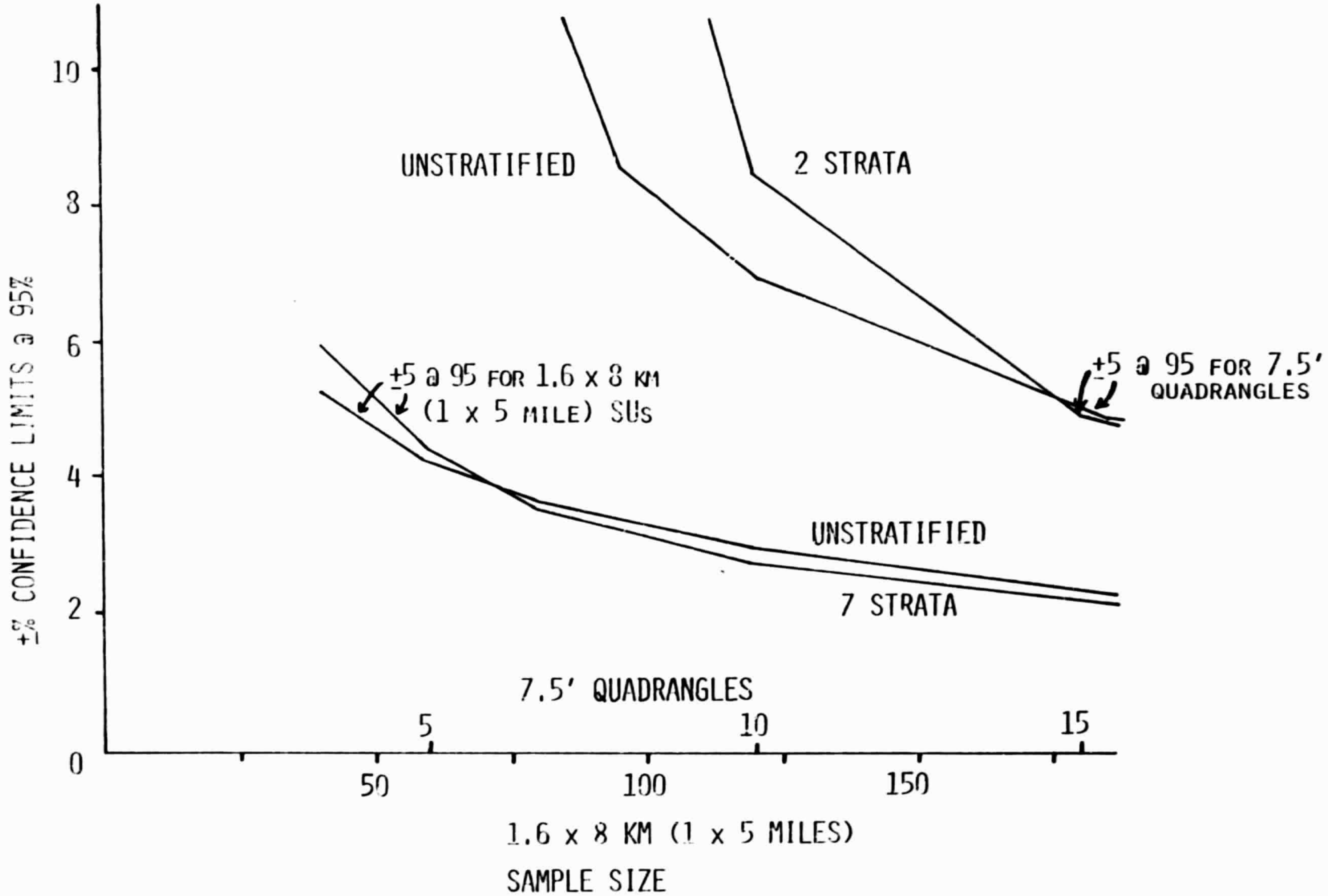


Figure 5-22. Comparison of accuracy of 1 x 5 mile sample units from Task I to 7.5' quadrangles used in Task II. The X axis is scaled so that the total area is the same for both sizes of sample units (e.g. 10 7.5' quadrangles ≅ 120 1 x 5 SUs ≅ 600 sq. mi.)

To assess the mapping accuracy of irrigated land in the Sacramento Valley 1° block, 32 of the 7.5' quadrangles were selected from the population of 64 in a checkerboard pattern. For each quadrangle, 24 points were systematically chosen (4 x 6 grid). The dot identified a field on which the following data was summarized: (1) irrigated or non-irrigated on the class map (Figure 5-20 a and b); (2) irrigated or non-irrigated on DWR's land use survey (Figure 1-2); (3) land use (crop type) assignment on DWR's land use survey; and (4) field size (< 4.05 ha, 4.05-7.69 ha, 7.70-15.78 ha, ≥ 15.79 ha [< 10 A, 10-19 A, 20-39 A, ≥ 40 A]).

For the 768 points over all 32 quadrangles the map accuracy was very good: percent correct = 94.0, percent omission = 7.4, and percent commission = 6.3. Examination of the four 30' blocks separately showed some deviation from the overall results (Table 5-5).

Table 5-5. Task II Map Accuracy

	% Correct	% Omission	% Commission
1° Block	94.0	7.4	6.3
30' Blocks			
Chico	94.3	3.3	13.0
Corning	96.4	4.8	6.3
Maxwell	96.9	7.1	1.5
Sutter	88.5	10.4	5.1

A closer examination showed that the errors were dependent on the percent irrigated in any particular 7.5' quadrangle (Figure 5-23). For quadrangles with low percent irrigated (< 33%) few errors of any kind occured. For moderate percent irrigated (33-67%) errors were primarily, but not exclusively, errors of commission. For high percent irrigated (> 67%) errors were primarily, but not exclusively, errors of omission. This pattern was significant using a Chi-squared test (p = 0.0003, Table 5-6).

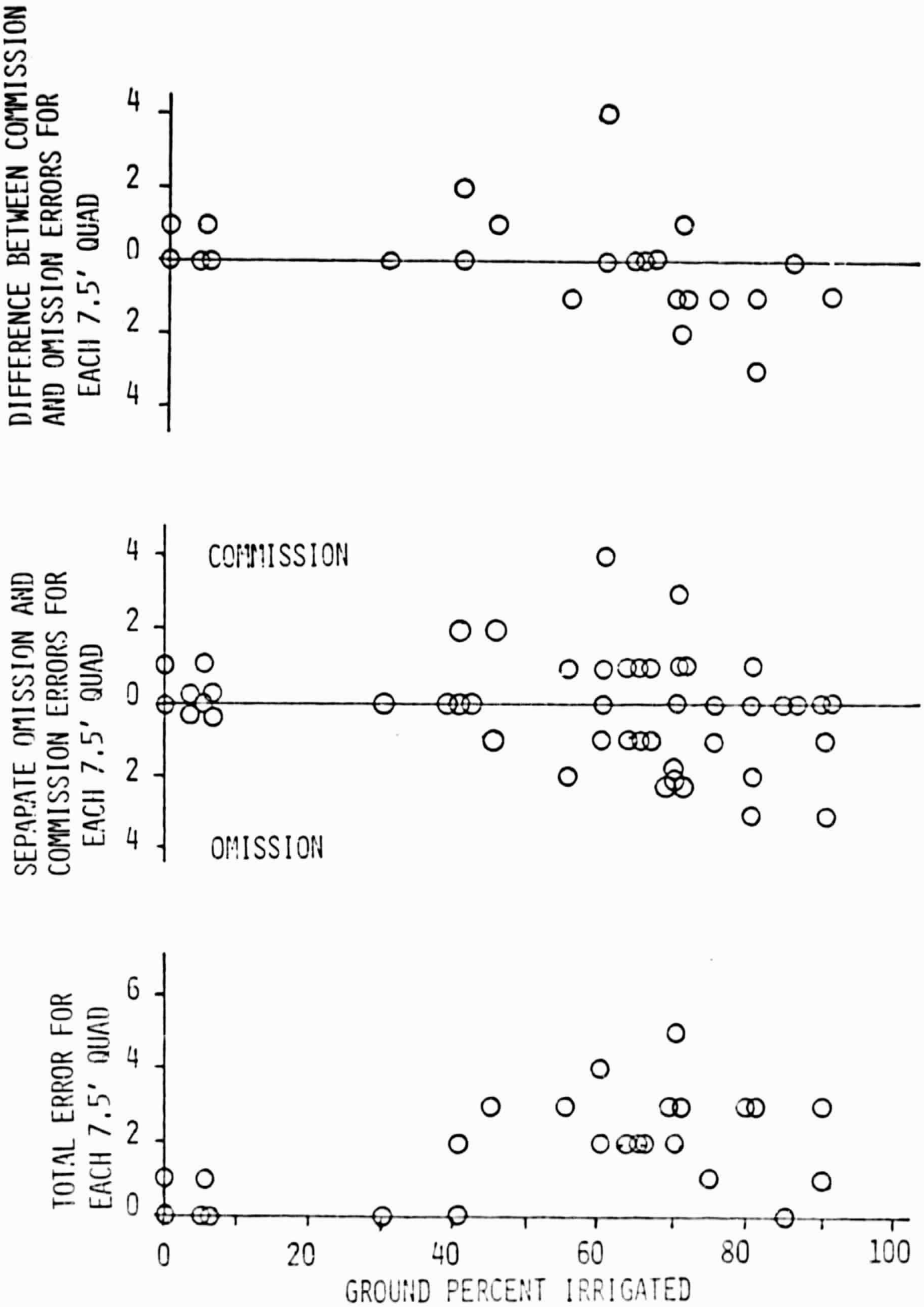


Figure 5-23. Relationship between percent irrigated and rates of commission and omission errors.

Table 5-6. Observed pattern of omission and commission errors as a function of percent irrigated. The value of Chi-square is for a test of the null hypothesis that omission and commission error rates are independent of percent irrigated.

	PERCENT IRRIGATED $\leq$ 67%	> 67%
OMISSION ERROR	7	18
COMMISSION ERROR	15	6
$\chi^2 \cong 8.8$, df = 1		
$p \cong 0.003$		

Many omission errors were in irrigated grain fields (13 of 25 points). Commission errors were generally in areas of native vegetation (13 of 21 points).

A simultaneous analysis of percent irrigated and crop/land use type showed no significant pattern of commission or omission error, i.e., omission errors were not primarily associated with grain located in areas of high irrigation; commission errors were not primarily associated with native vegetation located in areas of moderate irrigation.

Field size did have an effect on error rate. Proportionally more errors were made in small fields than in large fields. This pattern was significant using Fisher's exact probability test ($p \cong 0.011$, Table 5-7).

Table 5-7. Observed pattern of fields correctly and incorrectly classified as a function of field size for the 7.5' quadrangles.

	FIELD SIZE $\geq$ 16.19 ha (40 A)	< 16.19 ha (40 A)
CORRECT	500	20
ERROR	27	5

Continuing Work

Further analysis of the Sacramento Valley digital data set is planned for 1980. To continue the analysis an upgrade of UC Berkeley's Survey Planning Model (SPM) will be completed to allow inexpensive simulation of sample frame and irrigated (or crop) proportion(s) by spectral class over very large areas. The SPM will also allow simultaneous summary of irrigated proportion(s) by sampling stratum, measurement errors strata and reporting unit strata. Additionally, the SPM can be used to compute multivariate sample allocation for additional sample designs including ratio and regression. A test of the SPM on the 1976 Sacramento 1° block area will include the computation of first and second stage sample unit population variances for varying sizes of (1) primary sample units (PSU) (i.e., 7.5' quadrangles [155 x 200 cells], flight line strips [vertical 20 x 200 cells, horizontal 155 x 25 cells]) and (2) secondary sample units (SSU) (i.e., flight line strips, a field or field groups with a 7.5' quadrangle PSU). The test will also allow us to compute hypothetical PSU and SSU sample sizes and allocation among strata that minimize total variable cost subject to meeting pre-specified sample error requirements for an estimate of irrigated proportion.

5.3 PROPOSED WORK FOR 1980

The encouraging results of this digital analysis task make a large scale demonstration of the 7/5 ratio technique appropriate. Therefore, our objective for Task II in 1980 will be to define and demonstrate a Landsat, digitally-aided approach to estimating and mapping irrigated land on a hydrologic basin basis. Tentative demonstration areas are the Sacramento Valley and Tulare hydrologic basins.

For each of these areas, three dates of registered 1979 Landsat 7/5 data will be used for Phase I measurement. Setting of the 7/5 irrigation line (threshold) for creation of the class map will be done on interactive display and analysis systems available at the University. Sample unit data obtained for the Task I inventory will provide ground data.

Within this large scale demonstration several key design and evaluation activities will take place. (1) Definition of the procedure for setting an accurate Landsat 7/5 line needs to be refined. (2) Evaluation of map accuracy on a point-by-point basis by irrigation line threshold, region and combination of dates should be examined. (3) The form (linear, non-linear) and correlation of the Landsat to ground irrigated area relationship needs to be determined. (4) Cost and throughput rates should be documented, and (5) the Survey Planning Model could be used to determine (a) sample frame characteristics giving the lowest total variable cost (TVC) subject to error goals, (b) the impact of Landsat classification error on the final error of the irrigated proportion estimate, (c) the type of sample unit selection procedure that minimizes TVC subject to error goals and (d) the expected TVC, sample size and allocation among sample stages and strata that are necessary to meet given inventory error goals.

6.0 CROP TYPE ANALYSIS (TASKS III AND IV)

In the past, Task III (manual analysis) and Task IV (digital analysis) have been studied separately. In 1979, they were generally combined into one task with three major sub-tasks. The first sub-task began work on defining and understanding the nature of the complex and dynamic agricultural environment in California. The second sub-task began establishing a spatial, temporal, spectral base for a continuing (and increasing) effort in 1980. Third, the basic system for defining DWR's multicrop information requirements was started.

Three test sites were used for the multicrop analysis this year. UCSB studied a two-7.5' quad area on the Oxnard Plain of Ventura County (south coastal environment) and the three-quad area in Kern County described in Section 5.0. UCB concentrated work on the 1° block in the Sacramento Valley (also described in 5.0). See Figure 3-1.

6.1 WORK COMPLETED BY UCSB

6.1.1 Multicropping Studies

Largely as a result of California's generally mild climate, many areas support two or more crops per field in a single season. This results in an important temporal component in the design of a remote sensing program to monitor irrigated croplands. The timing of image acquisition and supporting field work is critical to the identification of specific crops and, to a lesser extent, irrigated land.

Multicropping Questionnaire

To determine the level of multicropping, the crops involved and the critical times, a questionnaire was sent to each county's agricultural extension office. Thirty-eight counties responded. The questionnaire was designed to gather information on the total acreage involved in multicropping as well as the specific crops and their planting sequences. An additional question probed those factors important in the farmer's decision to multicrop.

The results indicate that the northern portion of the state does not do much multicropping (generally less than 5% of the total farmed acreage), with that which does exist consisting primarily of small grains as the first crop. These are generally followed by pasture crops, sorghum or milo. In the southern and central interior portion of the state, considerably more multicropping occurs [Imperial County recorded over 40,500 hectares (100,000 acres) or approximately 25% of total farmed acreage in 1977]. While the small grains/sorghum or milo combinations are common, more vegetables are found in the multicropping sequence. The increased importance of multicropping in these areas is related to the warmer climate and availability of irrigation water. In the coastal areas (e.g. Salinas Valley and Oxnard Plain), multicropping tends to be widely practiced (over 75% of the land in truck crops is cropped more than once in a given season in Coastal Ventura County, according to local sources) with some fields

PRECEDING PAGE BLANK NOT FILMED

supporting three and four different truck crops per year. In these areas, a year-round mild climate, and high land rents combine to create dynamic cropping patterns. Such patterns require greater reliance on multidate imagery. The variety and intensity of vegetable multicropping in many of the coastal sites, where unique phenologies are not seen for many crops, may require the definition of crop groups.

Oxnard Plain Multicropping Dynamics Study

In order to determine the data requirements for monitoring both irrigated lands and crop type in a multicropping environment, a four date cropping dynamics study was undertaken on the Oxnard Plain. Early reports indicated that the combination of mild climate, excellent soils and high land rents resulted in intensive multicropping, with many fields yielding three and four crops in a single year. Our field work indicated that except for certain stable crops -- citrus, strawberries, turf grass and flowers -- crop rotation in this area occurs over short periods of time.

Four visits were made to the study site in 1979 (April, June, August and November). With the exception of the last visit, each data collection effort was 2 months apart. On each date, two field crews mapped crop type and field boundaries on a high altitude aerial photo base (1:32,500 scale). Additional information was gathered on crop growth state (emergent, young, mature) and conditions of particular interest (agricultural land being converted to urban use, removal of stable citrus groves for truck crop farming, etc.). The data was designed to document the type and location of crops in the area as well as the turnover rate for non-stable crops. Data was gathered for most of the agricultural lands on the Oxnard and Camarillo 7.5 USGS quadrangles.

In order to analyze the data, certain simplifications were required. A sampling scheme using a dot grid with 0.5 inches between each point was employed. The dot grid spacing was chosen on the basis of data manageability rather than for strict statistical reasons. Each dot represented approximately 10 hectares (25 acres). There were a total of 1464 dots for which there was field data on at least one of the four dates.

Table 6-1 shows the crop types or field conditions seen on each of the four dates. The major crop types are lemons, strawberries, sod and assorted vegetables. Of additional importance is the fairly large proportion of fallow cropland. In most cases, this category represents fields that are in preparation for planting. It is not clear if our field visits merely coincided with this phenomenon or if a large proportion of the fields can be expected to be fallow at any one time.

To evaluate the dynamics of multicropping two sets of tests were conducted -- three date and a two date analyses. For the three date tests (Table 6-2), those fields for which data were available on three consecutive dates were examined. For the April-June-August sequence, there were 538 fields; for June-

Table 6-1

Crops and Field Conditions Sampled on the Oxnard Plain

Crops	April	June	August	November
Fallow	326	250	249	357
Artichokes		1		
Asparagus	1	2		
Cole Crops	26	27	24	215
Celery	106	39	21	234
Lettuce	69	7	7	39
Melons/Squash		16	39	
Peas	16		5	
Spinach	27	16	1	
Tomatoes		144	198	21
Strawberries	42	36	42	60
Peppers		28	72	16
Parsley	7		13	17
Misc. Truck Crops	10	26	12	29
Corn		1	1	
Beans (dry)		219	410	
Flowers	21	6	4	4
Sod	21	29	40	47
Lemons	127	163	183	198
Oranges	2			
Avocado	3	8		

August-November, there were 928. Each field was then examined to determine the nature of the cropping sequence. Four characteristic sequences were noted -- the same crop or field condition on three different dates; the same crop or field condition on two consecutive dates; the same crop or condition on the first and third date, separated by a different crop on the second date; and, three different crops over the three date sequence. Table 6-2 shows the results for all three date sequences. In this case, fallow cropland conditions are treated as a crop-type.

If those fields where fallow conditions were noted on at least one of the dates are excluded, there is a reduction of more than 50 percent of the fields (Table 6-2). Nevertheless, those cases where different crops are found on each different date are a significant part of the total. If the fallow lands and stable areas (citrus and turfgrass primarily) are ignored, this condition, representative of intensive multicropping, represents between 35 and 45 percent of fields measured here.

The two-date test was designed to determine the amount of multicropping that could be expected over a two-month period. Three two-date sequences (April-June, June-August and August-November) were examined. For the April-June test there were 565 fields; June-August, 986; August-November, 1186. The results were tallied with and without fallow field conditions. The results are shown in Table 6-3. For each two date sequence (with or without fallow land), the condition of different crops being seen on different dates predominates.

The results of this sub-study indicate that in the type of environment characterized by the Oxnard Plain, where multicropping is commonly practiced, a significant number of air photo or Landsat acquisitions would be required to accurately determine irrigated acreage. Certainly the inadequacy of the current DWR procedure is evident here. Classification of particular crop types would probably be difficult for most of the truck crops because of their short time in the field and the fact that many, such as celery and lettuce, are found on all four dates and do not appear to have a unique phenological cycle. The definition of satisfactory crop groupings may alleviate some of these problems. Perhaps the principal value of Landsat in this situation is the ability to define intensive multicropping practices.

Our original research plan for the Oxnard Plain test site included a 1979 multidate Landsat analysis. A review of Landsat imagery available for 1979 indicates, however, that only one date of imagery was acquired during the primary growing season (April-September). This appears to be due to Landsat data processing problems rather than cloud coverage, since imagery was apparently not acquired for several clear day overpasses.

Table 6-3

Oxnard Plain Multicropping Dynamics

Analysis of 2 Consecutive Dates

Including Fallow Land

Date	XX	XY	Total
April/June	148 (26.2%)	417 (73.8%)	565
June/August	302 (30.6%)	684 (69.4%)	986
August/November	295 (24.9%)	891 (75.1%)	1186

Excluding Fallow Land

Date	XX	XY	Fallow
April/June	102 (18.1%)	129 (22.8%)	334 (59.1%)
June/August	245 (24.9%)	360 (36.5%)	381 (38.6%)
August/November	241 (20.3%)	442 (37.3%)	503 (42.4%)

6.1.2 Crop Phenology

Two crop phenology projects have been undertaken during this reporting period. The first involves the expansion of our crop phenology diagram series to include alfalfa, melons, and sugarbeets, in addition to cotton and small grains, previously completed. The second task involves a crop phenology survey of the Central Valley under the direction of Dr. Michael Nuttonson, former director of the Crop Ecology Institute. Dr. Nuttonson is designing a survey form and collecting necessary collatoral data. The survey will be conducted in cooperation with the University's Agricultural Extension Service.

6.1.3 Digital Crop Identification

Digital crop identification tests have been initiated in both Tulare and Kern Counties. As mentioned earlier, tests were scheduled for Ventura County (Oxnard Plain site) but suitable imagery has apparently not been acquired. Nine dates during the 1978 crop season have been obtained for Tulare County and reformatted into VICAR format. This data set will now be registered using the cross correlation techniques discussed earlier and a limited amount of DWR's ground truth data will be digitized and registered to the data set. This will allow automated labeling and performance evaluations.

As an alternative to the Ventura data set, we have begun crop identification tests using the five date Kern County data for 1976. Field crop maps acquired by a water district have been digitized for a three - 7 1/2 minute quadrangle site and registered to the Landsat data, as discussed earlier in our Task II section. Both supervised and unsupervised (clustering) approaches are being explored for the crop identification tasks.

6.2 WORK COMPLETED BY UCB

Working in the same Sacramento Valley 1° test site discussed in Section 5.0 UCB addressed two general issues. First, basic spectral/temporal data is needed on the major crops of the area as input for inventory design and classification procedures. Second, DWR ultimately requires output in a map-like form (preferably 7.5' quad) with the capability of recombining the data shown on the map in a variety of ways (i.e., by water district, study area).

The first step in pursuing spectral/temporal pattern of agriculture in this area was to determine the major crops and their spatial distribution. Using County Agricultural Commissioner's reports and DWR's 7.5' quads and tabulated summaries, a general description of agriculture in the whole Sacramento Valley (14 counties) and a detailed analysis of the five counties covered by the 1° block was done. For each of these five counties (Butte, Colusa, Glenn, Sutter and Tehama) reported crop acreages were tabulated (Table 6-4a). Crops were selected for specific study if they represented either 5% of any single county's

a

CROPS - 1976	BUTTE	COLUSA	GLENN	SUTTER	TEHAMA	TOTAL	% TOTAL	
•BARLEY	12,000	18,600	3,500	19,000	7,100	60,200	5.1	
BEANS	8,100	9,375	2,552	11,576	-	31,603	2.7	
•CORN	8,800	16,000	8,000	15,592	1,220	49,612	4.2	
•HAY, ALFALFA	6,500	3,440	16,000	9,715	4,000	39,655	3.4	
GRAIN	3,000	1,200	-	7,174	2,900	14,274	1.2	6.8
OTHER	1,150	-	3,000	20,430	2,000	26,580	2.2	
•OATS	5,000	-	-	1,043	2,500	8,543	.7	
•PASTURE, IRR.	19,800	12,000	36,000	24,000	32,200	124,000	10.5	
•RICE	70,000	108,000	53,149	78,964	-	310,000	26.2	
SAFFLOWER	-	8,100	888	8,226	-	17,254	1.5	
•SORGHUM	11,200	9,150	6,500	23,177	1,850	51,877	4.4	
SUGAR BEETS	4,040	12,800	7,222	5,648	815	30,525	2.6	
•WHEAT	27,600	39,000	22,500	40,000	14,000	143,100	12.1	
•FRUIT & NUT CROPS	64,976	22,150	21,062	46,238	25,663	180,139	15.2	
SEED CROPS	20,108	6,145	5,613	19,301	3,877	55,044	4.7	
VEGETABLE CROPS							3.4	
MELONS	-	-	-	1,556	-	1,556		
PUMPKINS	-	-	-	1,265	-	1,265		
SQUASH	-	-	-	322	-	322		
•TOMATOES, CANNING	-	8,000	-	24,500	-	32,500		
FRESH	-	-	-	85	-	85		
CORN, SWEET	-	-	-	178	-	178		
WATERMELONS	-	-	-	215	-	215		
MISCELLANEOUS	1,932	-	1,350	211	95	3,588		

b

	BUTTE	COLUSA	GLENN	SUTTER	TEHAMA	ALL
BARLEY	-	•	-	•	•	•
CORN		•		-		-
RICE	•	•	•	•		•
WHEAT	•	•	•	•	•	•
HAY			•	•	•	•
PASTURE	•		•	•		•
SORGHUM				•		-
FRUIT & NUTS	•	•	•	•	•	•
TOMATOES				•		

• = CROP REPRESENTED 5% OR GREATER OF REPORTED ACREAGE

- = CROP REPRESENTED BETWEEN 4 AND 5% OF REPORTED ACREAGE

GROUPING OF CROPS FOR TASK IV ANALYSIS:

RICE
SMALL GRAINS (WHEAT AND BARLEY)
ORCHARDS (FRUIT & NUT)
PASTURE (PASTURE AND HAY)
SORGHUM
CORN
TOMATOES

Table 6-4. Part "a" lists the acreages of crops in the five counties located in the 1°
block. Part "b" summarizes those crops that represented approximately 5
of the tabulated acreage.

total or, when combined, 5% of all the counties' acreage. Table 6-4b shows the crops which met or marginally met the 5% limit. The nine crops listed on Table 6-4b were further combined to the crop groups shown at the bottom of the figure.

It was felt that a sample of the 64 possible 7.5' quads in the area would provide ample statistics for our use. Using the 1976 DWR ground data, each 7.5' quad in the 1° block was examined for the presence of agriculture. Quads with less than 5% of their total area in cultivation were eliminated. The remaining 38 quads were divided into blocks of three and the one of every three with the most even distribution of the major crops was selected for analysis (Figure 6-1).

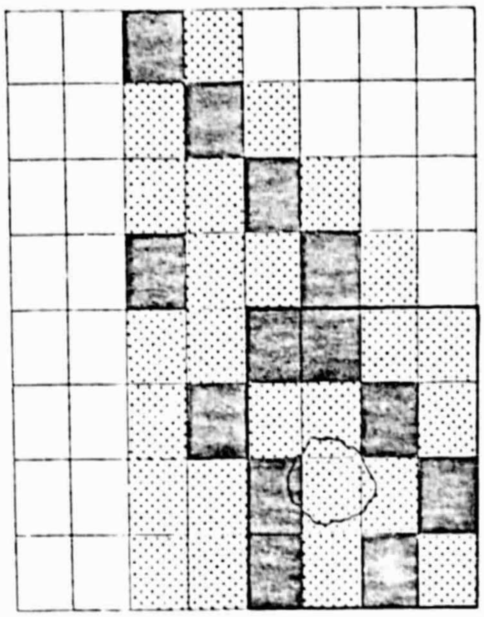


Figure 6-1. This diagram of the 64 7.5' quads in the 1° Sacramento Valley block shows (1) quads having less than 5% agriculture (no shade or pattern); (2) quads having 5% or greater agriculture that were not chosen for analysis (dotted pattern); and (3) quads having 5% or greater agriculture that were chosen for analysis (gray). Quads with the most even distribution of crops (based on DWR's land use survey summaries) were selected. The Sutter 30' block is located in the southeast corner and outlined with a heavy line (the irregularly shaped polygon is the perimeter of the Sutter Buttes).

Statistical data was obtained for the major crops in these selected 7.5' quads. Fields greater than 10 pixels, eliminating border pixels and field anomalies, were sampled for: (1) the mean 7/5 ratio value, (2) standard deviation, and (3) range of values. These statistics were tabulated by 7.5' quad, by 30' block and by 1° block for each of the three dates (May 30, August 28, October 4).

These data were examined for crop separability. Although these dates were selected for differentiation between irrigated and non-irrigated land, some crop types and groups were separable. Both small grains and rice have 7/5 values on these dates that allow them to be identified with little confusion. However, pasture and orchard appear similar, as do corn and sorghum. To spectrally separate these crops requires additional Landsat acquisitions.

The Sutter 30' block was selected for further analysis because of the high proportion of agriculture and the availability of additional dates of digital data. Two additional dates were chosen for analysis. First, May 4 for additional input to differentiate (1) small grains from native grasses and (2) pasture from orchard. Second, June 26 was chosen to (1) separate corn, with its earlier emergence, from sorghum and (2) identify tomatoes. The same 7.5' quads and sampled fields were used in this analysis. In addition to the 7/5 ratio band, a 5/4 ratio band and a sun angle corrected Euclidean brightness band were created for each of the 5 dates. For each of the ratioed bands and brightness band, statistics were combined over the 30' block by crop giving the mean value, standard deviation, and covariance for the five dates. These statistics were used to seed the unsupervised classifier (CLUSTER) on the RSRP interactive system.

A subset of dates was chosen to facilitate processing. On examination of the statistics and crop calendars, the May 4, June 26, and August 28 dates were chosen as giving the maximum separability between crops (Figure 6-2). Preliminary classification was done on every fourth pixel for six iterations giving a maximum of 30 classes and these results were used to seed the final clustering and labeling.

The combined statistics for the 30' block were used to label each of the 30 clusters. The statistics for the 7/5, 5/4, and brightness bands for each crop were compared to the mean values per band for each cluster (Table 6-5a). The clusters were given tentative labels as either: (1) one of the major crop types or (2) as other (Table 6-5b). Clusters with the same potential label were grouped and mapped as one color on the RSRP interactive display system. This display was checked against the DWR ground data maps (Figure 6-3). Visual comparison of the resulting output was encouraging, although no detailed statistical evaluation of the results has been done to date.

The function of much of the crop type tasks has been to provide baseline information and output products for the Department of Water Resources' evaluation. This basic data provides the background necessary to begin defining DWR's information requirements in reference to the use of a Landsat-based system. Information needs for both inventory and mapping systems must be carefully outlined for further work to proceed logically and efficiently.

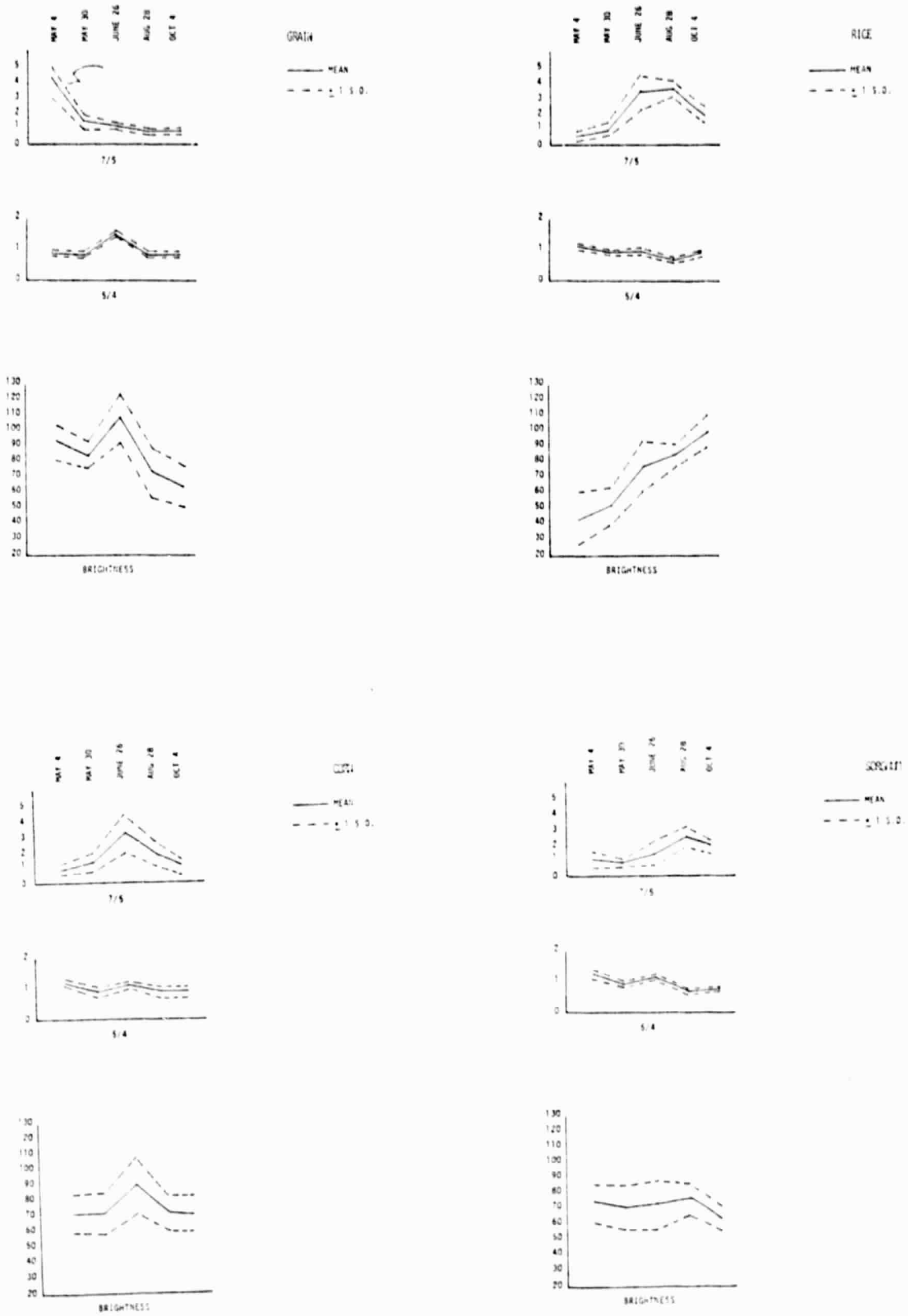


Figure 6-2. 7/5 ratio and 5/4 ratio and Euclidean brightness graphed against the five dates studied over the Sutter 30' block. The seven major crops found in the Sutter area are shown.

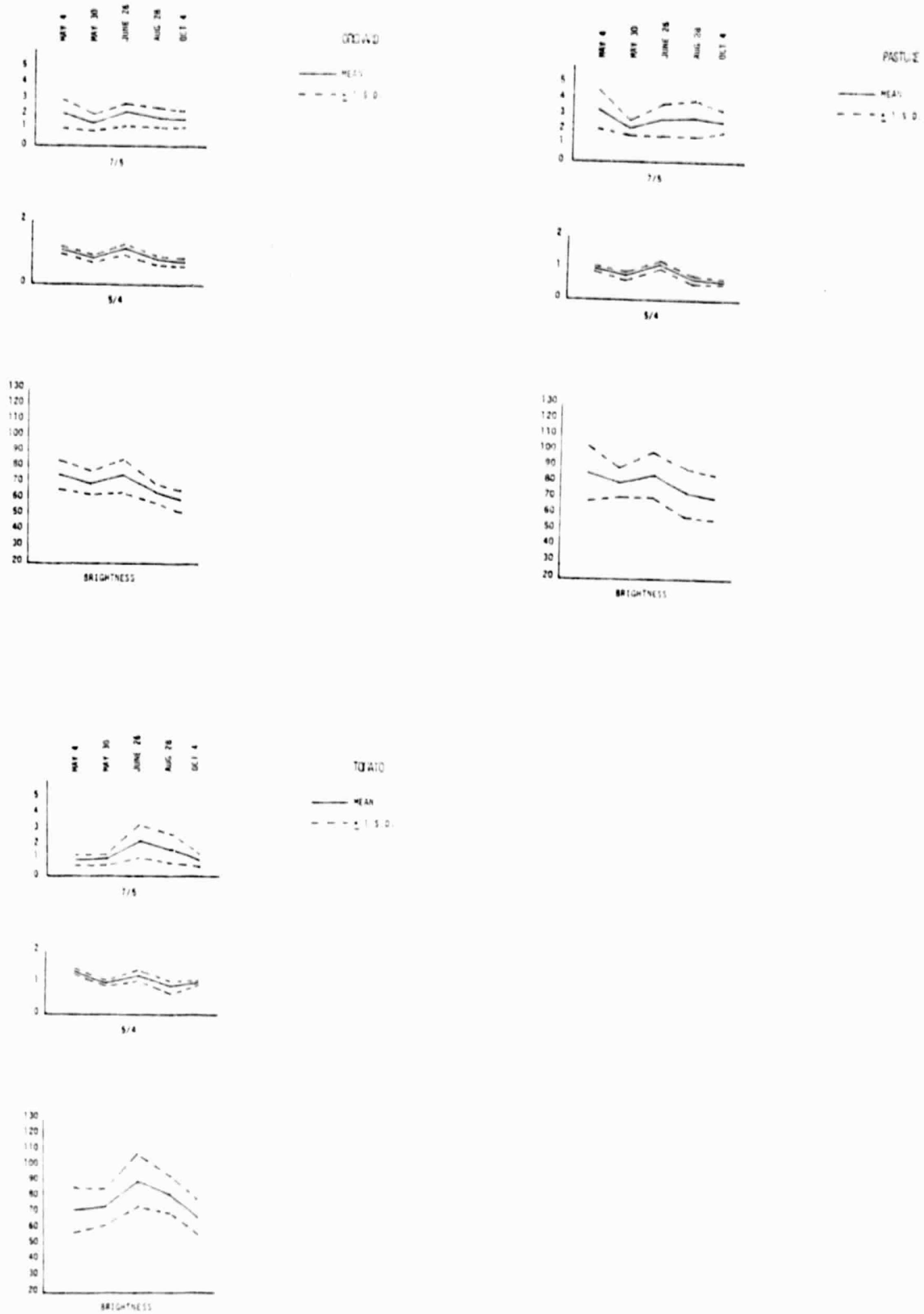


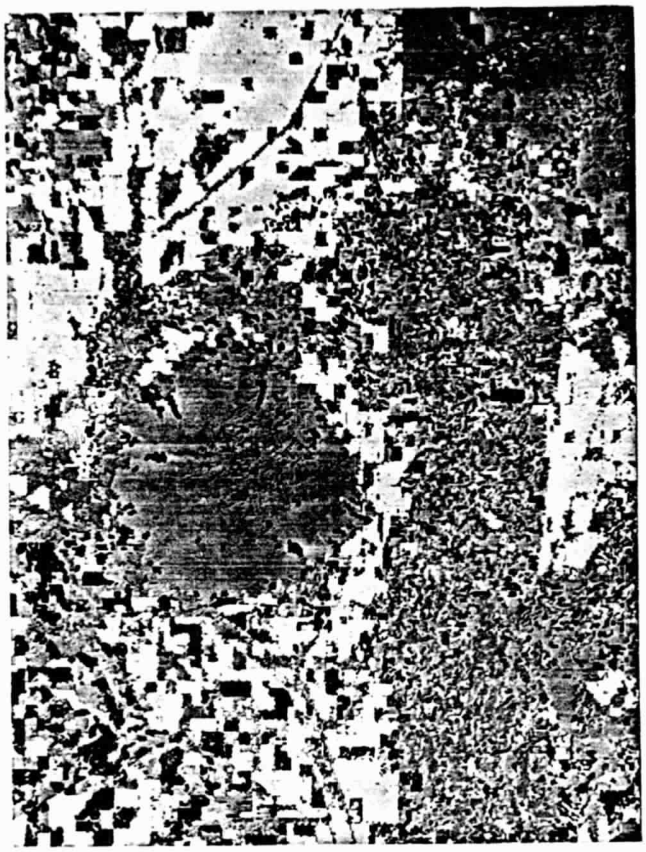
Figure 6-2 continued

Table 6-5a. Mean values of 7/5, 5/4 and brightness for each cluster.

CLUSTER	Aug 7/5 CH(1)	May 7/5 CH(2)	June 5/4 CH(3)	Aug 5/4 CH(4)	June 7/5 CH(5)	May 5/4 CH(6)	May Brightness CH(7)	June Brightness CH(8)	August Brightness CH(9)
1	[illegible]	2.24	1.55	[illegible]	1.19	1.06	[illegible]	110.41	98.52
2	[illegible]	.93	.89	.70	3.56	[illegible]	[illegible]	81.74	82.59
3	1.76	1.12	1.68	.56	[illegible]	1.25	[illegible]	94.16	[illegible]
4	[illegible]	.98	1.34	.46	[illegible]	1.36	[illegible]	32.59	[illegible]
5	.86	1.42	1.55	1.13	1.10	1.36	[illegible]	103.22	[illegible]
6	[illegible]	4.25	1.06	.48	3.22	.94	[illegible]	100.43	85.72
7	1.62	1.76	1.27	.88	1.59	1.21	[illegible]	83.83	71.28
8	[illegible]	1.59	.93	.67	4.13	1.23	[illegible]	[illegible]	95.61
9	3.26	.46	.95	.63	2.47	1.08	[illegible]	64.59	83.38
10	[illegible]	.94	.97	.61	1.74	1.32	57.87	54.63	80.21
11	2.15	1.04	1.28	.74	1.24	[illegible]	[illegible]	[illegible]	77.75
12	.98	1.35	1.46	1.02	1.16	1.35	[illegible]	83.50	73.83
13	[illegible]	3.57	1.27	.84	1.77	[illegible]	97.69	92.92	74.91
14	[illegible]	2.67	1.30	[illegible]	1.46	[illegible]	[illegible]	75.18	50.50
15	[illegible]	4.79	1.64	.95	1.19	[illegible]	97.74	117.45	[illegible]
16	[illegible]	.32	[illegible]	.43	4.08	[illegible]	29.68	[illegible]	87.75
17	[illegible]	1.11	.92	.83	4.18	[illegible]	[illegible]	104.55	67.81
18	[illegible]	.70	.76	[illegible]	4.22	[illegible]	[illegible]	[illegible]	83.62
19	[illegible]	1.18	1.11	.79	2.37	[illegible]	[illegible]	[illegible]	80.66
20	[illegible]	[illegible]	1.00	.47	3.34	[illegible]	[illegible]	[illegible]	[illegible]
21	[illegible]	[illegible]	1.20	.48	.72	1.01	[illegible]	[illegible]	40.17
22	[illegible]	[illegible]	1.53	[illegible]	1.26	1.14	[illegible]	[illegible]	73.59
23	[illegible]	[illegible]	1.22	[illegible]	.58	[illegible]	[illegible]	53.79	107.45
24	[illegible]	[illegible]	1.13	.72	2.17	[illegible]	[illegible]	[illegible]	[illegible]
25	[illegible]	[illegible]	1.14	.75	[illegible]	[illegible]	[illegible]	97.53	[illegible]
26	[illegible]	[illegible]	1.15	[illegible]	[illegible]	.86	[illegible]	[illegible]	76.16
27	[illegible]	[illegible]	1.46	[illegible]	1.37	[illegible]	[illegible]	[illegible]	[illegible]
28	[illegible]	[illegible]	1.30	[illegible]	[illegible]	1.22	[illegible]	69.97	[illegible]
29	[illegible]	[illegible]	[illegible]	.72	1.77	1.14	47.45	63.59	[illegible]
30	[illegible]	2.43	[illegible]	[illegible]	1.41	[illegible]	[illegible]	75.12	[illegible]

Table 6-5b. Cluster labels based on values shown above.

CLUSTER	CLUSTER
1 - grain	16 - rice
2 - rice	17 - rice
3 - corn	18 - rice
4 - sorghum	19 - other
5 - tomato and other	20 - rice
6 - pasture	21 - other
7 - orchard	22 - other
8 - other	23 - other
9 - rice	24 - orchard
10 - rice	25 - rice
11 - sorghum	26 - grain/sorghum double cropped
12 - other	27 - other
13 - pasture	28 - other
14 - orchard	29 - rice
15 - grain	30 - pasture



Blue = Small grains
Brown = Pasture
Gray = Other
Green = Sorghum
Orange = Rice
Red = Orchard
Yellow = Corn

Figure 6-3. Sutter 30 minute block. Classification of crop type based on cluster labeling.

To guide the development of a proper inventory system, certain key questions need to be addressed. Some of the questions are:

- What are the parameters for which estimates are desired?
  - Proportion of area by crop type?
  - Change in proportion by crop type?
  - Water demand by crop type?
- What are the target crops for which error will be controlled?
- What are the error goals for the target crops?
- Are there other crops or land use classes for which parameter estimates are desired?
- What are the target populations (areas) for which information is required?
- At what reporting unit are the data to be summarized?
- At which reporting level is error to be controlled?
- What are the constraints on the system?
  - Cost?
  - Timeliness?
  - Institutional capability?

Producing an accurate map requires responses to a different set of queries:

- What are the crop or land use classes to be mapped?
- What is the minimum acceptable classification accuracy by class?
- What is the maximum acceptable field boundary error?
- What are required characteristics of the map product?
- What are the constraints on the system?

Working closely with DWR, we anticipate specifying a set of inventory and mapping goals and constraints and proceeding with a larger scale demonstration in 1980.

Appendix I: A Comparison of Estimate Accuracy and Costs for Segment and Transect Sampling - UCSB

In preparation for this year's APT segment sampling, the cost of acquiring sample segments in a completely random manner using medium scale aerial photography appeared large and perhaps out of proportion to true statistical value. Although the large cost may be due to overly conservative estimates in terms of the number of segments that can be flown per day, it seems appropriate that more economical sampling schemes also be considered. An obvious alternative to random sampling designs which would make good use of photographic plane time is transect sampling along predetermined flight lines. The following analysis looks at this type of sampling scheme, which would fit well into DWR's present procedures. It is assumed that measurement error would remain the same regardless of sample design, so the two areas of concern are sample error and costs.

Our earliest work in this area was covered in the semiannual progress report of June, 1979. Using a similar approach to the one cited here, it was found that random segment and systematic transect samples, containing approximately the same amount of area, yielded estimates of irrigated acreage that were not significantly different from one another. A two-to-one cost differential for photo acquisition made transect sampling appear to be preferable to the segment approach. Subsequent to that effort, it was brought to our attention that the variance in the transect sample had not been computed properly (each transect had been treated as a group of 1.6 X 8 km (1 X 5 mile) segments rather than a single sample). The work presented here is a more thorough comparison of the two sampling approaches.

Comparison of Estimate Errors

Using a map of active cropland in Kern County for the 1978 growing season (Figure I-1), a test was conducted to determine the impact of segment and transect sampling on the estimate of cropland acreage. Data preparation involved the tabulation of the proportion of each square mile devoted to agriculture to the nearest 5 percent. The tabular data was placed into a computer data file so that each square mile could be accessed by its X-Y coordinate. This resulted in a data matrix of 62 X 79 elements representing 12,539 square kilometers (4898 square miles). However, because of a large amount of area which was not used for agricultural purposes, there were only 6902 square kilometers (2696 square miles) for which data was compiled. Analysis of the entire population of 2696 elements showed that the average amount of irrigated land was 58.2 percent with a standard deviation of 41.3 percent.

The test consisted of determining the number of segment and transect samples that must be taken before the estimate of cropland acres becomes stable. Stability in the estimate was determined by taking the standard deviation of the estimate over 100 iterations for each sample size.

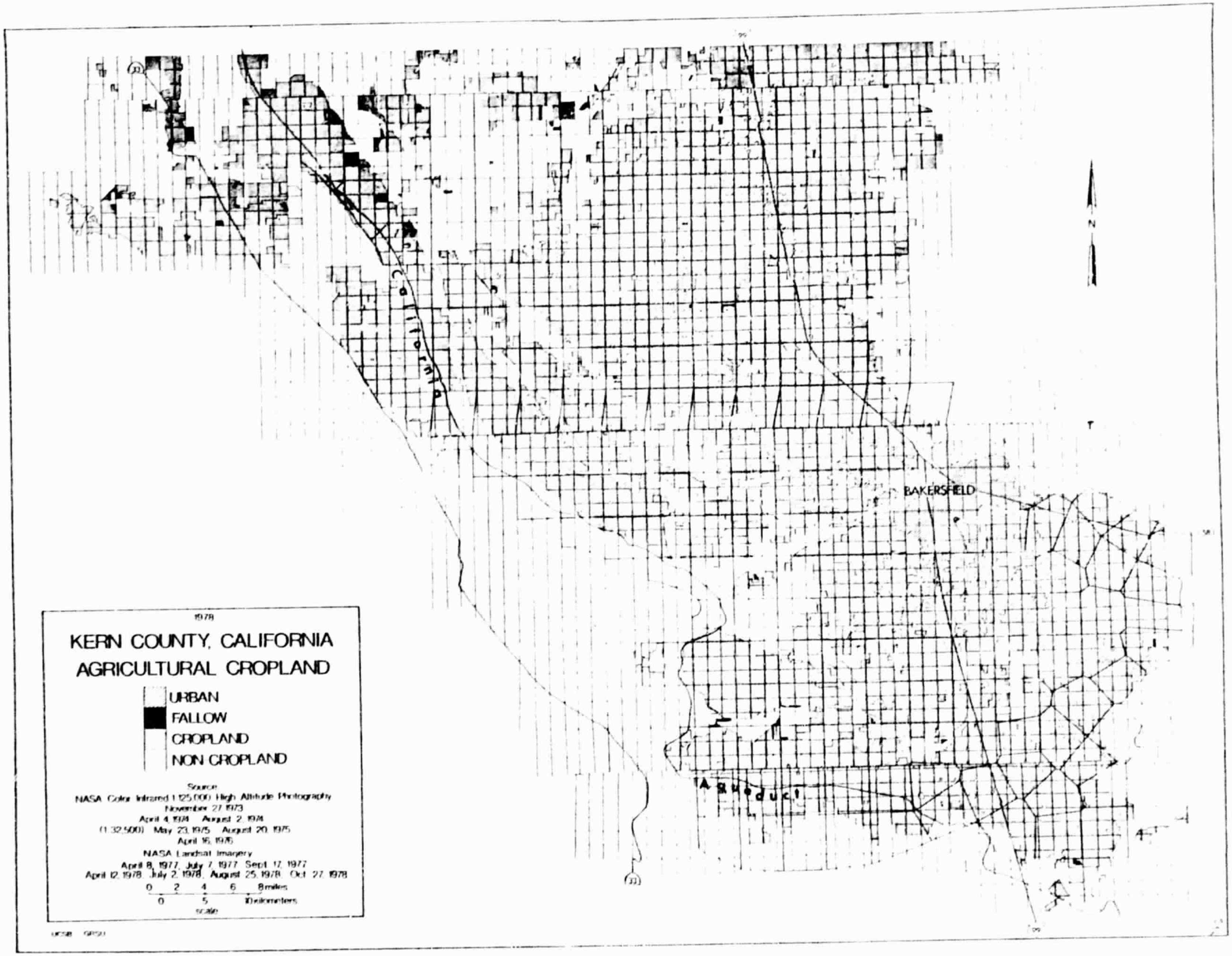

1978
KERN COUNTY, CALIFORNIA
AGRICULTURAL CROPLAND
URBAN
FALLOW
CROPLAND
NON CROPLAND
Source
NASA Color Infrared 1:125,000 High Altitude Photography
November 27, 1973
April 4, 1974 August 2, 1974
(1:32,500) May 23, 1975 August 20, 1975
April 16, 1976
NASA Landsat Imagery
April 8, 1977 July 7, 1977 Sept 17, 1977
April 12, 1978 July 2, 1978 August 25, 1978 Oct 27, 1978
0 2 4 6 8 miles
0 5 10 kilometers
scale
BAKERSFIELD
California
Aqueduct
N

Figure I-1

Based on earlier work documented in the ILP study, a single stratum (the Kern County study area was treated as such) of this size would require 40 segment samples. This would yield coverage of approximately 512 square kilometers (200 square miles). To test the behavior of the estimate, computations were made for sample sizes of from 20 to 70 segments (by 5 segment increments). In all cases sampling was random and without replacement. Table I-1 shows the average estimate derived from each sample size and the standard deviation of the estimate. For samples sizes of 60, the variance in the estimate is significantly reduced and remains relatively stable or reduces gradually for larger sample sizes (Figure I-2).

Random transect sampling was undertaken for a range that would result in approximately the same amount of area being covered as in the random segment test. One hundred iterations of transect sample sizes ranging from 3 to 10 were made. All transects were taken randomly and without replacement. The results are shown in Table I-2.

As can be seen in Figure I-3 for any given amount of area to be covered, the variance of the estimate is greater for the transect sample than the segment sample. This indicates that a greater amount of area must be flown to yield a stable and dependable estimate of irrigated acreage.

A test was also conducted to assess the impact of systematic sampling as opposed to total randomness. The map was divided into 8 fields of ten transects each. Using a sample size of 8 transects, the variance in the estimate over 100 iterations was computed. For an areal coverage of approximately 440 square kilometers an estimate of 58.41 percent was computed. The standard deviation of the estimate using systematic transect sampling was 4.2 percent, whereas random transect and random segment sampling yielded standard deviations of 6.2 percent and 4.9 percent, respectively. It appears that the distribution of agricultural land for this area is clumped such that a systematic type of approach results in a better estimate. The improved results seen here are analagous to improvements in sampling when stratification is used. While not tested here, it may be that a more systematic (or stratified) approach to random segment sampling would improve the performance of that procedure.

Cost Comparison

A comparison of costs was somewhat difficult because it required that we make assumptions about the time required to acquire the aerial photographs. Current estimates are that approximately 20 five-mile random sample segments can be flown per day. According to Fred Stumpf (DWR-San Joaquin District), Fresno County's 526,000 hectares (1.3 million acres) can be flown in 5 days using a transect approach. This is equivalent to 640 km (400 miles) per day.

Using the average area covered as a measure of total transect length (Table I-2) and a value of 126 km (79 miles) as the inter-transect distance (this) is the maximum width of the study area), the cost, in flight-days, at the rate of 640 km per day, was computed. This is shown in Figure I-3. Also seen

Table I-1

Random Segment Sampling Results

#Segments	Mean	Variance	Standard Deviation	km^2
20	58.88	50.45	7.10	160.30
25	58.95	39.09	6.25	200.60
30	58.97	26.01	5.10	240.90
35	58.44	34.50	5.87	281.62
40	56.44	26.04	5.10	325.12
45	57.96	24.14	4.91	374.85
50	59.74	23.64	4.86	402.76
55	60.44	25.45	5.05	443.56
60	58.84	17.28	4.16	480.27
65	59.34	14.40	3.79	523.49
70	58.91	15.42	3.93	570.10

Figure I-2 Standard Deviation of Estimate

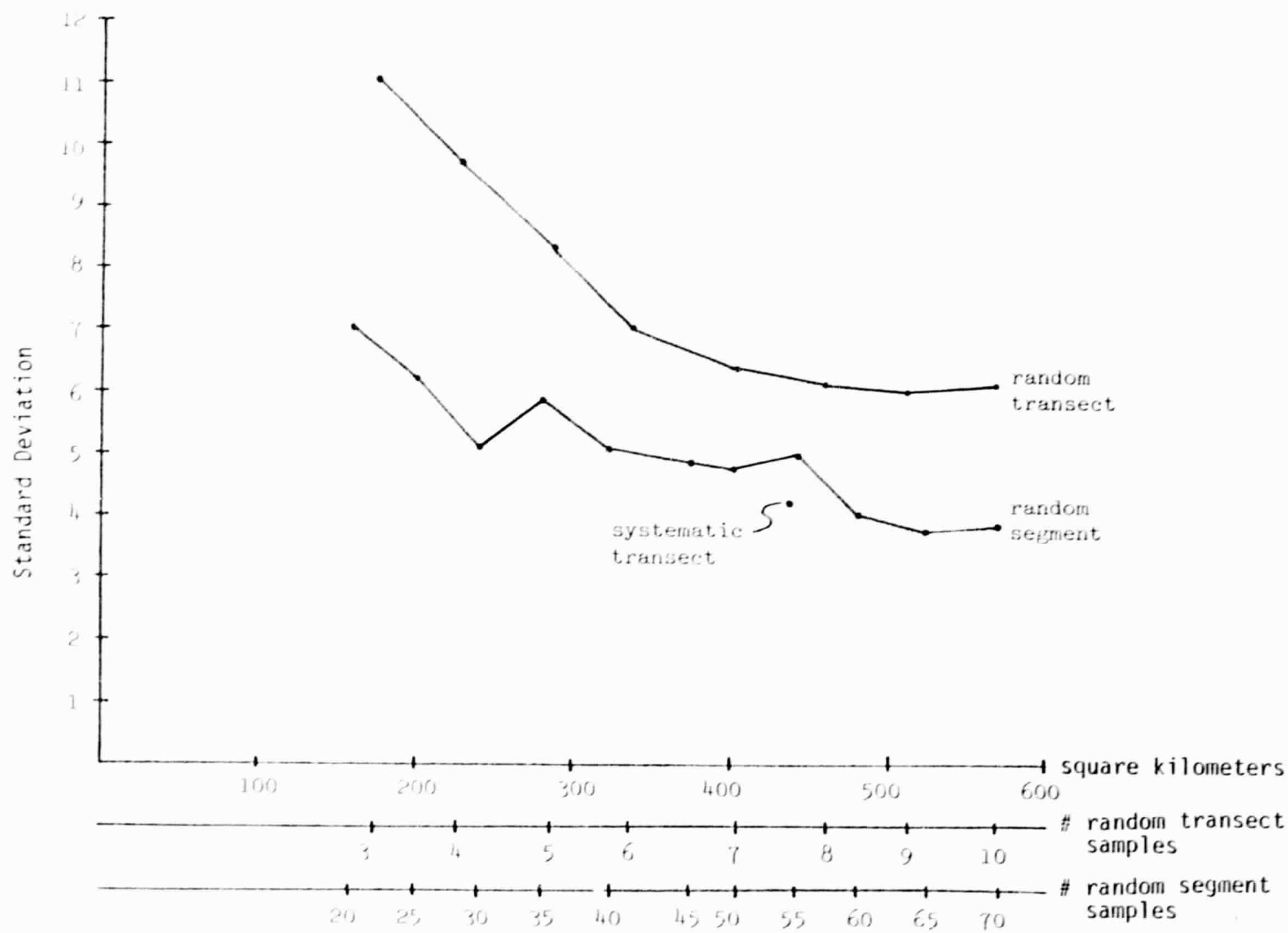
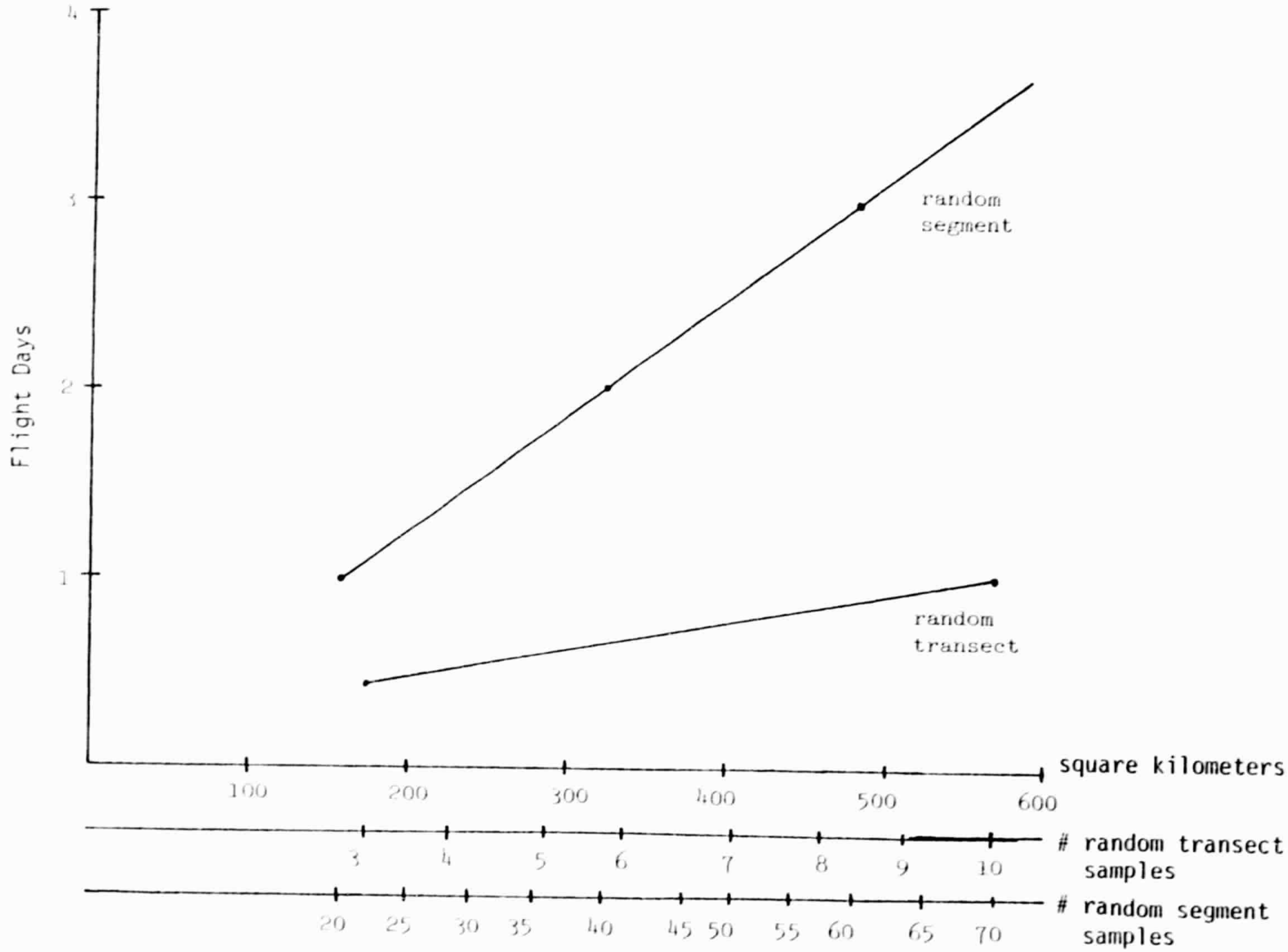


Table I-2

Random Transect Sampling Results

#Transects	Mean	Variance	Standard Deviation	km^2
1	56.40	440.65	20.99	61.26
2	56.10	216.65	14.72	116.44
3	59.10	125.96	11.22	175.80
4	57.18	95.23	9.76	228.80
5	57.10	70.67	8.41	285.25
6	56.41	50.39	7.10	336.88
7	57.21	42.05	6.48	405.79
8	58.22	39.58	6.29	461.53
9	57.87	37.59	6.13	513.18
10	57.53	39.38	6.28	568.46
11	58.31	28.17	5.31	633.89
12	58.21	26.03	5.10	696.50

Figure I-3 Flight Time Requirement for Sample Acquisition
(by Area Covered)

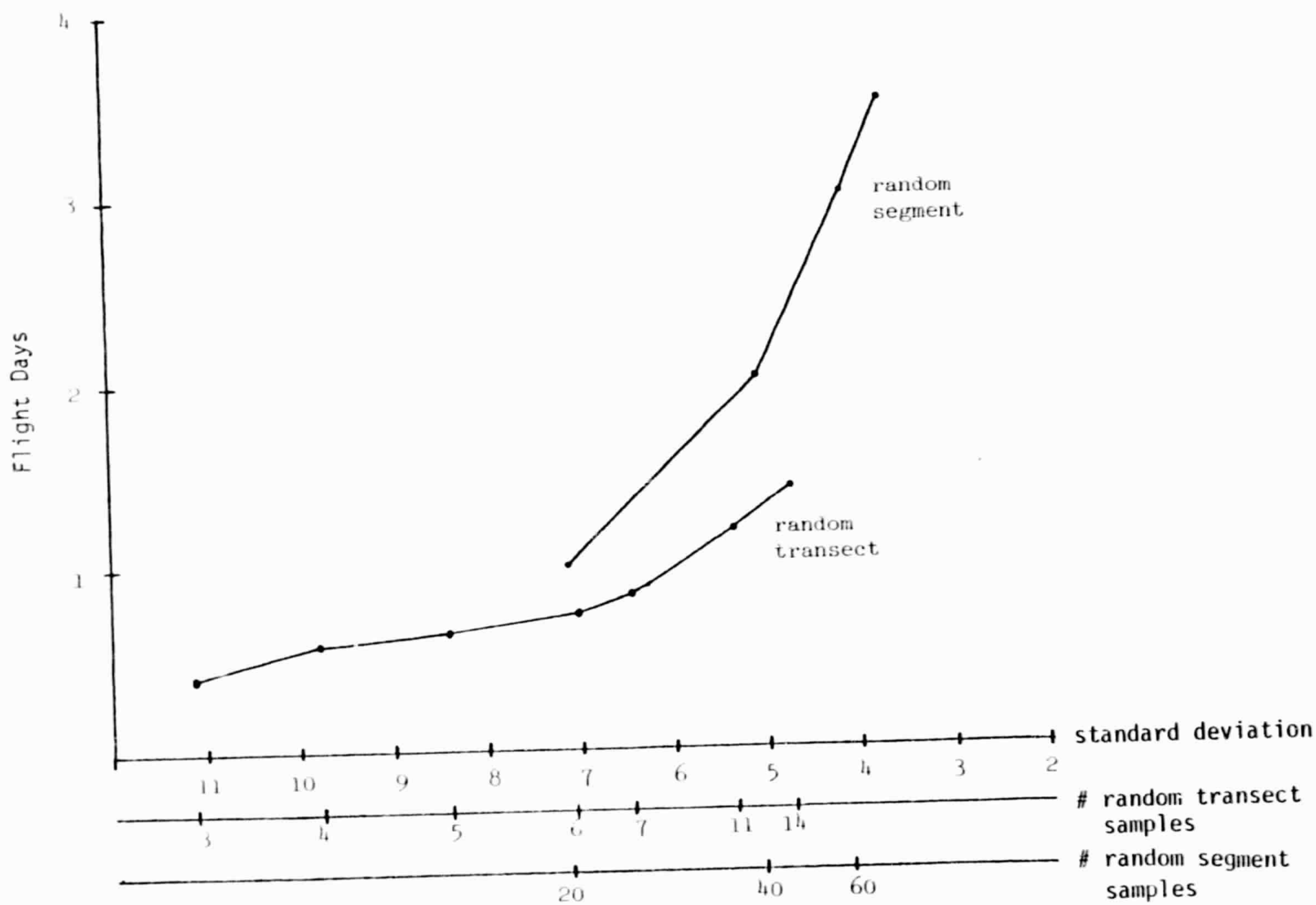


here is the number of flight-days required to gather random segments, at the rate of 20 per day. As can be seen in Figure I-3, flight time requirements are significantly greater for segment sampling, primarily as a result of increased inter-segment search time. The operational simplicity of the transect procedure results in lower acquisition costs. Because DWR's photo acquisition costs have been approximately $135 per hour in the past, flight time can be a siginificant part of a multistage sampling program employing aircraft.

Figure I-4 shows the cost in flight days of both the segment and transect approach for given levels of variance in the estimate. In all cases the random transect cost is lower than that for segment sampling. Furthermore, as greater reductions in variance are achieved through higher sampling rates the cost difference between the two techniques increases.

These results indicate that transect sampling may be a cost effective alternative to random segment sampling. The chief trade-off involved is the requirement for acquisition of greater photo coverage which increases the cost of the photo interpretation phase of the project.

Figure I-4 Flight Time Requirements for Sample Acquisition
(by Standard Deviation)



Appendix II: Accessibility Categories Used to Predict the Weighted Average Relative Cost (c_i)

Assuming that some sample units require more time to access and ground check, a relative measure of this different accessibility in terms of a cost ratio was needed for the optimal sample unit allocation process. For the present inventory, each sample unit was assigned one of three accessiblity types:

Type A: The sample unit was (1) near other sample units, (2) near a good access road, and (3) had a road network going through or by it.

Type B: The sample unit lacked one or two of the Type A requirements to some extent.

Type C: The sample unit lacked all three criteria of a Type A unit to some extent, or one criteria was completely lacking.

Within a county, each polygon defined by the merged stratification (Section 4.2.4) was evaluated using available USGS 1:250,000 scale topographic maps and assigned an accessibility code. Relative costs associated with each accessibility type were estimated from experience in the 10-County and 14-County studies and current DWR costs on the following parameters:

- The cost to photograph one sample unit using a 35 mm camera in a light aircraft is $32.
- In areas defined as Access Type A, field crews could reasonably be expected to ground check five sample units per day.
- In areas defined as Access Type B, four sample units per day.
- In areas defined as Access Type C, three sample units per day.
- The cost of maintaining a ground crew in the field is $220/day.

Using these figures, relative costs for ground data collection were estimated for each accessibility type. The relative costs shown in Table II-1 were used to predict the average relative cost (c_i) for each stratum.

Table II-1. Cost of collecting sample unit ground data, by accessibility type.

$$\text{Cost} = \frac{\text{Ground crew cost per day}}{\text{Number of sample units collected per day}} + \text{Cost of processing aerial photography per sample unit}$$

Accessibility Type	COST	Relative Cost c_i (COST/76)
A	$\frac{\$220}{5} + \$32 = \$76$	1.00
B	$\frac{\$220}{4} + \$32 = \$87$	1.14
C	$\frac{\$220}{3} + \$32 = \$105$	1.38

As these values for photo and ground data acquisition are general approximations, more refined estimates will be realized at the conclusion of the 1979 inventory for future operational sample allocation efforts. The DWR ground survey crews recorded time spent ground checking the sample units allowing these more refined estimates for future surveys.

Appendix III: Landsat Acquisitions Used for Task I (listed by county)

Alameda	47-34			
	2 June			
	26 July			
	18 Sept			
Alpine	46-33			
	8 April			
	3 Aug			
	26 Sept			
Amador	46-33	47-33		
	8 April	–		
	3 Aug	26 July		
	26 Sept	–		
Butte	47-32	47-33		
	11 June			
	26 July			
	18 Sept			
Calaveras	46-33	46-34		
	28 June	23 May		
	3 Aug	3 Aug		
	26 Sept	26 Sept		
Colusa	47-33	48-33		
	18 April	23 March	+ [illegible] June partial	
	26 July	18 July		
	18 Sept	14 Aug		
Contra Costa	47-34			
	2 June			
	26 July			
	18 Sept			
Del Norte	49-31	50-31		
	17 May	–		
	19 July	–		
	11 Sept	–		
El Dorado	46-33	47-33		
	8 April	2 June		
	3 Aug	26 July		
	26 Sept	18 Sept		
Fresno	45-34	45-35	46-34	46-35
	4 May	4 May	23 May	23 May
	6 July	7 Sept	3 Aug	3 Aug
	4 Oct	16 Sept	26 Sept	3 Aug
Glenn	48-32	48-33		
	28 April	23 March		
	18 July	18 July		
	28 Sept	14 Aug		
Humboldt	49-31	49-32		
	17 May	22 June		
	19 July	24 Aug		
	11 Sept	11 Sept		
Imperial	41-37	42-37		
	12 April	10 May		
	14 June	17 Aug		
		22 Sept		
Inyo	43-35	44-34	44-35	45-34
	5 April	15 April	15 April	4 May
	22 July	10 Aug	23 July	20 Aug
	25 Sept	6 Sept	30 Oct	4 Oct
Kern	44-35	44-36	45-36	[illegible]-36
	15 April	15 April	4 May	2 March
	23 July	23 July	7 Sept	6 July
	30 Oct	30 Oct	16 Sept	6 July
Kings	45-35			
	4 May			
	7 Sept			
	16 Sept			
Lake	48-32	48-33		
	28 April	23 March		
	18 July	18 July		
	28 Sept	14 Aug		
Lassen	47-31	47-32		
	2 June	2 June		
	26 July	26 July		
	18 Sept	18 Sept		
Los Angeles	44-36			
	15 April			
	23 July			
	12 Oct			
Madera	45-34	46-34		
	4 May	23 May		
	6 July	3 Aug		
	4 Oct	26 Sept		
Marin	47-34	48-33	48-34	
	–	23 March	28 April	
	–	18 July	18 July	
	18 Sept	14 Aug	–	
Mariposa	46-34			
	23 May			
	3 Aug			
	26 Sept			
Mendocino	49-32	49-33	48-33	
	22 June	6 March	23 March	
	24 Aug	24 Aug	5 Aug	
	11 Sept	(haze)	14 Aug	
Merced	46-34			
	23 May			
	3 Aug			
	26 Sept			

Modoc

47-31	48-31
2 June	2 June
11 June	~~#~~ 28 April
26 July	18 July
18 Sept	28 Sept

Mono

46-33	48-34	45-33	44-34
–	6 July	13 May	–
3 Aug	20 Aug	6 July	6 Sept
26 Sept	4 Oct	16 Sept	–

Monterey

46-35	47-35	47-34
23 May	4 March	11 June
3 Aug	26 July	18 Sept

Napa

48-33
23 March
18 July
14 Aug

Nevada

47-32	47-33	46-33
2 June	28 June	–
26 July	26 July	3 Aug
18 Sept	18 Sept	26 Sept

Orange

43-37	44-36
22 July	15 April
22 July	23 July
12 Oct	30 Oct

Placer

46-33	47-33
3 April	18 April
3 Aug	26 July
26 Sept	18 Sept
2 June	2 June

Plumas

47-32
11 June
26 July
18 Sept

Riverside

41-36	42-36	43-36	41-37
12 April	10 May	25 June	12 April
2 July	30 July	22 July	14 June
27 Oct	22 Sept		27 Oct

42-37	43-37
10 May	29 May
17 Aug	22 July

Sacramento

47-33	47-34
2 June	2 June
26 July	26 July
18 Sept	18 Sept

San Benito

46-34	46-35	47-_4
23 May	23 May	2 June
3 Aug	3 Aug	26 July
26 Sept	18 Sept	18 Sept

San Bernardino

41-36	42-36	43-36	42-35	43-35
12 April	25 May	29 May	10 May	5 April
2 July	30 July	22 July	30 July	22 July
3 Sept	22 Sept	23 Sept	22 Sept	23 Sept

San Diego

42-37	43-37
10 May	22 July
30 July	22 July
17 Aug	Sept

San Francisco

47-34
11 June
26 July
18 Sept

San Joaquin

47-33	47-34
11 June	11 June
26 July	26 July
18 Sept	18 Sept

San Luis Obispo

45-35	45-36	46-35	46-36
4 May	2 March	23 May	28 June
7 Sept	6 July	3 Aug	3 Aug
			20 Sept

San Mateo

47-34
11 June
26 July
18 Sept

Santa Barbara

45-36	46-36
2 March	18 April
6 July	28 June
6 July	26 Sept

Santa Clara

47-34
11 June
26 July
18 Sept

Santa Cruz	47-34		
	11 June		
	26 July		
	18 Sept		
Shasta	48-31	48-32	
	28 April	28 April	
	18 July	18 July	
	28 Sept	28 Sept	
Sierra	47-32	47-33	46-33
	2 June	2 June	2 June
	26 July	26 July	3 Aug
	18 Sept	18 Sept	26 Sept
Siskiyou	49-31	48-31	
	17 May	28 April	
	19 July	18 July	
	11 Sept	28 Sept	
Solano	47-33	47-34	
	2 June	2 June	
	26 July	26 July	
	18 Sept	18 Sept	
Sonoma	48-33		
	23 March		
	18 July		
	3 Aug		

Stanislaus	46-34	47-34	
	23 May	11 June	
	3 Aug	26 July	
	26 Sept	18 Sept	
Sutter	47-33		
	18 April		
	26 July		
	18 Sept		
Tehama	48-32		
	28 April		
	18 July		
	28 Sept		
Trinity	49-31	49-32	
	22 June	22 June	
	19 July	19 July	
	11 Sept	11 Sept	
Tulare	44-35	45-34	45-35
	15 April	4 May	4 May
	23 July	6 July	7 Sept
	30 Oct	4 Oct	16 Sept

Tuolumne	46-33	46-34
	8 April	8 April
	3 Aug	3 Aug
	26 Sept	26 Sept
Ventura	45-36	44-36
	2 March	15 April
	6 July	23 July
	6 July	30 Oct
Yolo	47-33	48-33
	18 April	23 March
	26 July	18 July
	18 Sept	3 Aug
Yuba	47-33	47-32
	18 April	2 June
	11 June	26 July
	18 Sept	18 Sept

Appendix IV. Publications Issued Under NASA Grant NSG-2207

1979

Tinney, L., J. Baggett and M. Cosentino. "A Multistage Mapping Approach to Agricultural Surveys Using Satellite and Aircraft Imagery." Proceedings Third Conference on the Economics of Remote Sensing (November 1979).

Tinney, L., J. Holloway, J. Baggett and J. Estes. "A Multistage Mapping Approach to Inventorying Irrigated Cropland Using Landsat and Aircraft Imagery." Proceedings 5th Pecora Symposium (July 1979).

Tinney, L., S. Wall, R. Colwell and J. Estes. "Irrigated Lands Assessment for Water Management - Applications Pilot Test." Proceedings 5th Pecora Symposium (July 1979).

Wall, S.L. "California's Irrigated Lands: Landsat-Based Estimation and Mapping." Proceedings Symposium on Identifying Irrigated Lands Using Remote Sensing Technologies (November 1979). Missouri River Basin Commission, Omaha, Nebraska.

Wall, S.L. and J. Baggett. NASA Grant NSG-2207 Quarterly Progress Reports for the periods:

1 December 1978 - 30 March 1979
1 April 1979 - 30 June 1979
1 July 1979 - 30 September 1979

Space Sciences Laboratory, University of California, Berkeley.

Wall, S.L., R.W. Thomas and L.R. Tinney. "Landsat-Based Multiphase Estimation of California's Irrigated Lands." Joint Proceedings of the ASP-ACSM 1979 Fall Technical Meeting (September 1979), 221-236.

1978

Wall, S.L., L. Tinney, J. Baggett, C.E. Brown, K.J. Dummer, T.W. Gossard, J. Holloway, T. Torburn and R.W. Thomas. "Irrigated Lands Assessment for Water Management - Applications Pilot Test (APT). Annual Progress Report: 1 November 1977 - 31 December 1978. Series 20, Issue 7. Space Sciences Laboratory, University of California, Berkeley.

Wall, S.L., L. Tinney, J. Holloway and T. Torburn. "Irrigated Lands Assessment for Water Management - Applications Systems Verification and Transfer (ASVT). Semi-Annual Progress Report: 1 November 1977 - 30 April 1978. Series 19, Issue 59. Space Sciences Laboratory, University of California, Berkeley.

Wall, S.L., L. Tinney, C.E. Brown, S.J. Daus, C.E. Ezra, T. Torburn and V.L. Vesterby. "Determining the Usefulness of Remote Sensing for Estimating Agricultural Water Demand in California. Annual Report 1 January 1977 - 28 February 1978. Space Sciences Laboratory, University of California, Berkeley.

Appendix IV (continued)

1977

Wall, S.L., L. Tinney, E. Aderhold, S.J. Daus, E. Ezra, T. Hardoin, M. Kronenberg and V.L. Vesterby. "Determining the Usefulness of Remote Sensing for Estimating Agricultural Water Demand in California. Semi-Annual Report: 1 January 1977 - 30 June 1977. Series 18, Issue 59. Space Sciences Laboratory, University of California, Berkeley.